AD 625262

SPEECH INTERFERENCE ASPECTS OF NAVY NOISES

Calculations show speech interference measurements are simplified by averaging sound pressure levels in mid-frequency octaves

J. C. Webster and R. G. Klumpp

Research and Development Report · 2 September 1965

U. S. NAVY ELECTRONICS LABORATORY, SAN DIEGO, CALIFORNIA 92152 · A BUREAU OF SHIPS LABORATORY

THE PROBLEM

Determine simple methods for rating noise in shipboard spaces in relation to its interference with speech communication.

RESULTS

1. Representative samples of ship, office, and shop noises were recorded, measured, and analyzed.

2. Naval ship spaces tend to be noisier than civilian spaces where equivalent communicating jobs are performed.

3. Sixteen noises were selected out and adjusted in intensity to be equally speech-interfering. Simple physical measurement and calculations on these 16 equally speech-interfering noises showed that Speech Interference (SI) could be measured:

a. best by a ^^aging the Sound Pressure Levels (SPL) in mid-frequenc_ octaves (300 to 600, 600 to 1200, 1200 to 2400 cycles per second (c/s)), called the Speech Interference Level (SIL) method;

b. next best by using weighting networks A or Din 3 in Sound Level Meters, or by finding the SIL (averaging the SPL's) in the octaves from 300 or 600 to 4800 c/s.

c. least well by fitting spectral noise peaks to Noise Criterion rating curve contours of which the Noise Criterion Alternate (NCA) was better than the conventional Noise Criterion (NC) or International Standards Organization (ISO) contours.

4. More complex physical measurement or calculation on the 16 equally speech-interfering noises showed that Speech Interference could be measured well by Articulation Index (AI) methods. Simpler 5- and 6-octave methods employing a generalized speech spectrum were almost as good as the more elaborate 20-band method using the actual speech spectrum utilized in this experiment.

5. Speech Interference could be predicted well by using families of NC, NCA, or ISO curves if:

a. only that part of any of the curves that centered at 500, 1000, and 2000 c/s were used;

b. the curves "averaged through" spectral peaks and valleys of the noise spectra.

6. For speech in quiet, half the intelligibility lies in frequencies above and below some value at or between 1600 and 1900 c/s. As the ratio of speech to noise intensity deteriorates, the frequency that divides the speech spectrum into two equal halves (as regards contribution to speech intelligibility) drops from 1600 or 1900 c/s to about 800 or 1000 c/s.

7. A new Speech Interference (SI) noise rating contour was developed that could be used in any of the conventional ways of measuring Speech Interference, namely:

a. to estimate the SIL;

b. as a weighting network for a Sound Level Meter;

c. as a noise-rating contour.

8. The new SI contours rated the Speech Interference effects of the 16 noises as good or better than any previous method, and in addition resolved many of the extreme differences among the three speech interference rating methods.

9. The new Speech Interference Contours are not drastically different from theoretical extensions of the AI calculation method.

10. Thermal noises (TN) with spectra shapes of -12, -6, and +6 dB per octave (TN-12, TN-6, TN Flat, and TN+6) are representative of the steady-state noises in the original 16 noises.

11. The 16 equally speech-interfering noises were neither equally loud nor equally annoying.

12. Maximum noise level for face-to-face communication is a 500/1000/2000-c/s SIL of 95 dB.

13. Maximum 500/1000/2000 SIL for speech communication when using good "noise-proofed," sound-powered-phones is:

a. 84 dB if the talker is in the quiet and the listener is in noise;

b. 94 dB if both the talker and listener are in noise;

c. 114 dB if the talker is in noise, the listener in quiet.

14. Amplified speech communication with earphones is possible in a 500/1000/2000-c/s SIL of 120 dB if use is made of noise-cancelling dynamic or condenser microphones, noise shielding at mouth and ear, a speech bandwidth of three octaves or greater centered between 1000 and 1800 c/s, a low sidetone level, AVC, and peak clipping.

15. Amplified speech communication with loudspeakers is possible in 500/1000/2000-c/s SIL noise levels of 80 dB. If earplugs or passive earmuffs are worn, this level can be extended to 95 dB.

RECOMMENDATIONS

1. In using SIL methods to rate noises, average the sound pressure levels in the octaves centered at 500, 1000, and 2000 c/s. This procedure is a compromise between the presently used 600 to 4800-c/s range and the range found to be best in this study, the 300 to 2400-c/s range.

2. Use the newly developed SI contours as:

   a. a new weighting method for Sound Level Meters;

   b. extensions of the existing noise-rating contours.

3. Use of the SI contours should be evaluated in working ships' spaces.

4. Loudness and annoyance aspects of noises should be considered in future extensions of this work.

5. Spaces where conversations cannot be carried on in comfort at 3 feet are too noisy for tasks requiring face-to-face communications. In general this is when the average noise level in the octaves, centering at 500, 1000, and 2000 c/s (the 500/1000/2000-c/s SIL), exceeds 70 dB.

6. If the 500/1000/2000-c/s SIL exceeds 90 dB, the wearing of hearing protection should be mandatory.

ADMINISTRATIVE INFORMATION

Work was performed under SF 013 11 01, Task 1357 (NEL N3-4) by members of the Listening Division, as assigned by BUSHIPS ltr ser 345-438 of 3 January 1961. The report covers work from January 1961 through December 1964 and was approved for publication 2 September 1965.

The assistance of J. L. Leonard, a coauthor of Section IV, who performed many of the physical measurements on the equally speech-interfering noises; D. D. Washburn and T. H. Wells who assisted in the collection, analysis, and editing of the shipboard noises; W. E. Green, who aided in running the experiment to adjust the levels of noise to make them equally speech-interfering; and R. S. Gales and R. W. Young, who critically read and offered constructive suggestions on this report, is gratefully acknowledged. To these, and to the ships' personnel who assisted in the noise collection phase of the study, sincere thanks are extended.

PREFACE

This report consists of seven reprints of papers in professional journals plus collateral and supplementary material. Because of this there will be some duplication, and the historical and/or logical order will not always be followed strictly.

The problem in terms of a real naval environmental control problem is stated in Section I (Introduction). The evidence that the problem really exists is presented in Section II (a summary of noise surveys aboard a number of ships). Sections III, IV, and V are reprints of papers dealing with two experiments on the physical and speech-interfering properties of diverse spectrum noises.

Section VI gives some details on psychophysical measurement methods for noises.[1*] Section VII, which has been submitted for publication, and Sections VIII and IX (reprints), propose new ways of measuring the speech-interfering properties of noise.

Section X shows the important frequency regions in noise-masked speech in terms of where in the speech impairment-handicap-disability scale the criterion is chosen.

Section XI summarizes speech capabilities in noise and is based primarily on evaluations of the speech intelligibility of numerous communication systems and components.

---

* See list of references at end of report.

CONTENTS

Contents (Continued)

ILLUSTRATIONS

Illustrations (Continued)

TABLES

INTRODUCTION

The Bureau of Ships has specifications[2,3] limiting the levels of noise in various ship spaces and for emission of noise by equipment. These specifications are based on considerations of the effects of noise on communications and on the potential deafening effects.

This report reexamines the question of speech interference effects of noise with a view toward simplifying the noise level measurement procedures. To accomplish these tasks: (1) noise levels in a number of shipboard spaces were measured aboard operating ships (and reviewed from other shipboard measurements) (Section II); (2) representative aircraft, ship, machinery, and office noises were collected (Section III); (3) 12 of these noises plus four laboratory-generated noises were then spectrum-analyzed (Section III); (4) intelligibility tests were conducted to find the levels at which these 16 noises interfered equally with speech (Section III); (5) the 16 noises were subjected to a number of simple analytical physical tests and measurements to find a physical measurement which agreed that the noises were equally speech-interfering (Section III); (6) the variability among estimations of the average noise level obtained from observers reading moving coil meters was determined (Section IV); (7) complex calculation schemes based on physical measures were applied to the 16 noises (Section IV); (8) psychophysical measurement schemes were applied (Section VI); and (9) a new set of speech Interference Criteria was developed (Sections VII, VIII, and IX).

On the basis of this reexamination, two summary papers were compiled that concern the important frequencies in noise-masked speech (Section X) and noise limitations on speech (Section XI).

Related studies were conducted to determine the relationships between (1) noise levels and speech levels,[4] and (2) the angle between talkers and listeners in face-to-face communications.[5,6] These are presented in Appendixes A, B, and C.

REVERSE SIDE BLANK

MEASUREMENT OF SHIP NOISES

To obtain experience with the kinds of levels of noise aboard ships, noise levels were measured on three Navy vessels, the aircraft carriers USS ORISKANY (CVA 34) and USS TICONDEROGA (CVA 14), and the missile cruiser USS CANBERRA (CAG 2). Measurements were made during sea trips taken in conjunction with two other problems dealing with communications in the Combat Information Center (CIC) and on the flight deck of aircraft carriers. To some degree, therefore, the sampling tended to be concentrated in areas occupied by CIC and Air Department personnel.

Figures II-1 and II-2 summarize, in histogram form, 157 noise levels measured aboard the three ships. Except for a few large compartments, each measurement represents a single compartment. The uppermost histogram of figure II-1 shows the distribution of overall or C scale levels. All measurements were made with calibrated General Radio Company Type 1551-A or 1551-B Sound Level Meters. The arrow at 86 dB indicates the median level. The next histogram depicts sound levels obtained with the A weighting [7] of the sound-level meter for the same set of 157 measurements. For this distribution the median is 76 dB.

The two lower histograms, figure II-2, present overall (C scale) and A-weighting levels measured in spaces in which satisfactory speech communication was judged by the measurement team to be important. The 64 measurements selected from the 157 measurements correspond roughly to the A category of Ships Specification SI-10.[2] The median C and A levels for these spaces are 82 and 70 dB, respectively.

Figure II-3 presents histograms of noise levels measured by Jensen and Soroka aboard the aircraft carrier USS CORAL SEA (CVA 43).[8] The upper histogram is based on overall levels calculated from listed octave band levels and the lower histogram is based on Speech Interference Levels[9] of the same measurements. Median levels are 93 and 73 for the two distributions of 60 measurements each.

Additional data on noise levels aboard U. S. Navy ships are available in reports from the Material Laboratory, New York Naval Shipyard.[10] Median overall levels were about 84 dB for 44 measurements on USS BORIE (DD 704), about 90 dB for 20 measurements on USS TIMMERMAN (EAG 152), and about 80 dB for USS TICONDEROGA for 19 measurements taken with no aircraft in operation.

Speech Interference Levels (the 300 to 4800-c/s, four-band average) gave median values of 80 dB for 20 measurements on USS TIMMERMAN and of 66 dB for 20 measurements on USS TICONDEROGA.

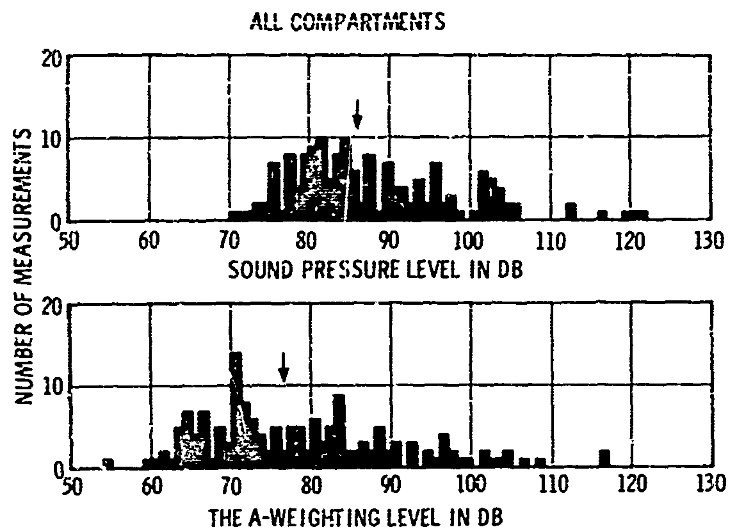


Figure II-1. Distribution of noise levels in representative compartments on two aircraft carriers and one heavy cruiser. The upper histogram is measured with the C-, the lower with the A-weighting, of a sound level meter. Arrows point to median values.

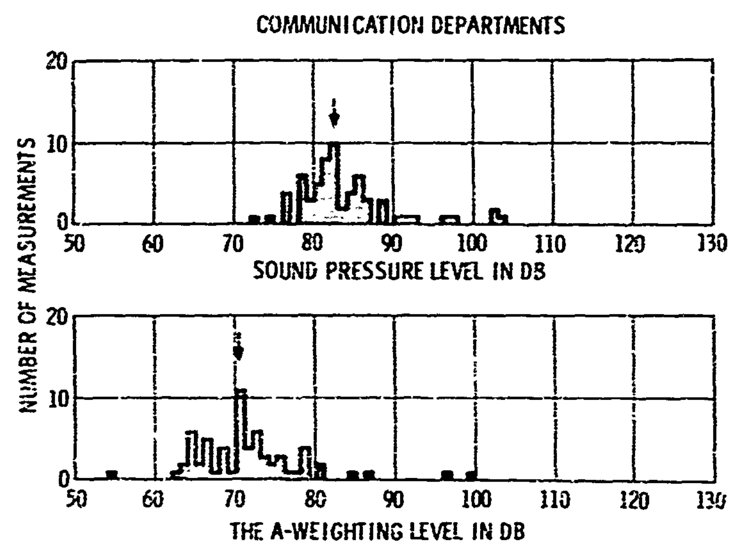


Figure II-2. Distribution of noise levels in only those compartments where speech communications were judged to be necessary on two aircraft carriers and one heavy cruiser. The upper histogram is measured with the C-, the lower with the A-weighting, of a sound level meter. Arrows point to median values.

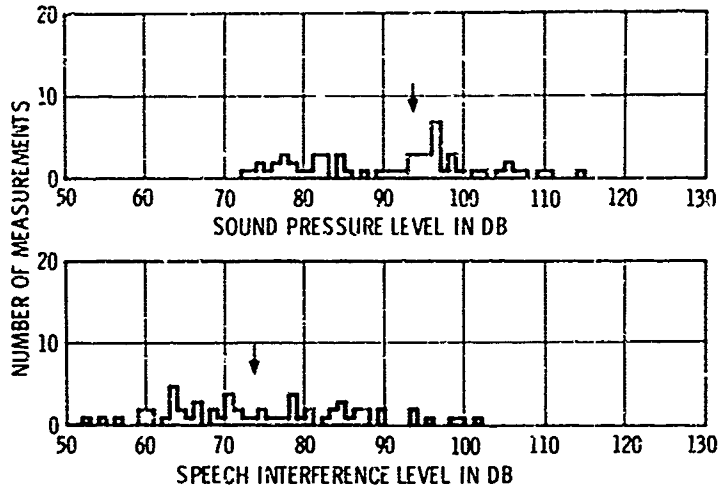


Figure II-3. Distribution of noise levels aboard USS CORAL SEA (CVA 43). The upper histogram represents C levels, the lower is the average level in the four octaves from 300 to 4800 c/s, the four-octave Speech Interference Level (SIL). Arrows point to median values.

A later survey of shipboard noises by the Southwest Research Institute[11] concludes "... that airborne sound levels generated by machinery items are above specifications in several shipboard spaces."

Figures II-4 through II-11 summarize the SRI noise measurements and the BUSHIPS specifications applicable at the time. The measurements reported were made aboard the aircraft carriers USS ENTERPRISE (CVA 65), USS FORRESTAL (CVA 59), USS KITTY HAWK (CVA 63), USS RANGER (CVA 61), and USS HANCOCK (CVA 19) from 1959 to 1962; on various conventional and nuclear submarines in 1958 and 1963; and destroyers in 1956 and 1963. The authors measured noise levels only aboard DD's 849 and 868 and CVS 18 in March 1964. In general the noises in Category D and E spaces were measured at full power runs and other spaces during "endurance" running conditions.

The noise levels aboard merchant ships of the Netherlands[12] and Norway[13,14] varied from 65 dB(A) or 95 dB(C) in cabins to 105 dB(A) or 110 dB(C) in Engine Rooms.

Noise levels on the navigation (pilot) bridge of 24 German ships[15] varied between 70 and 102 dB at 31 c/s and between 40 and 55 dB at 2000 c/s, with spectra falling about 10 dB per octave from 31 to 250 c/s and about 5 dB per octave above 250 c/s. On the average, C levels were 90 dB at "full speed ahead" (volle Fahrt) and 80 dB at "stop."

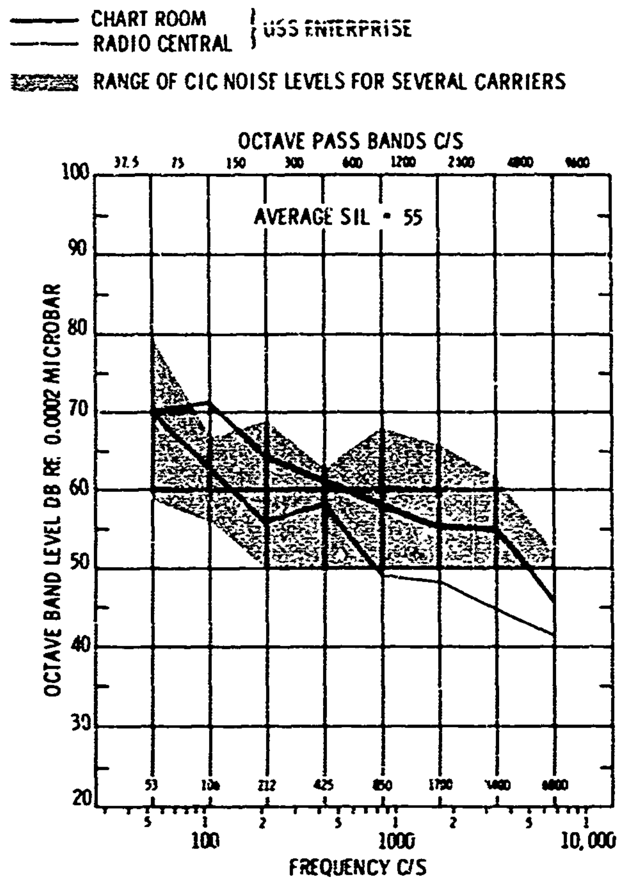


Figure II-4. The range for three Category A spaces (aircraft carriers). SIL limits: CIC - 50 dB; Chart Room and Radio Control - 60 dB.

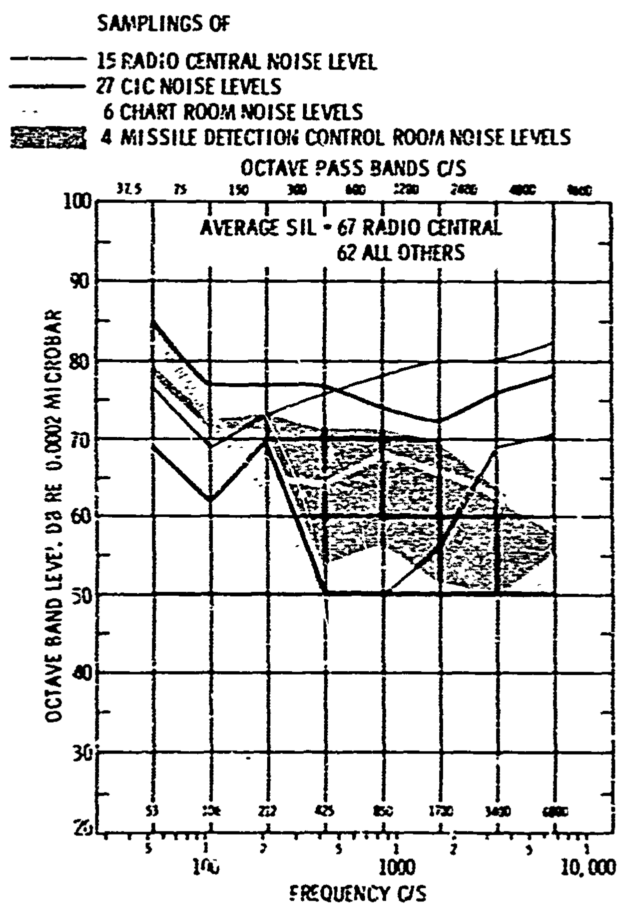


Figure II-5. The range for four Category A spaces (destroyers).

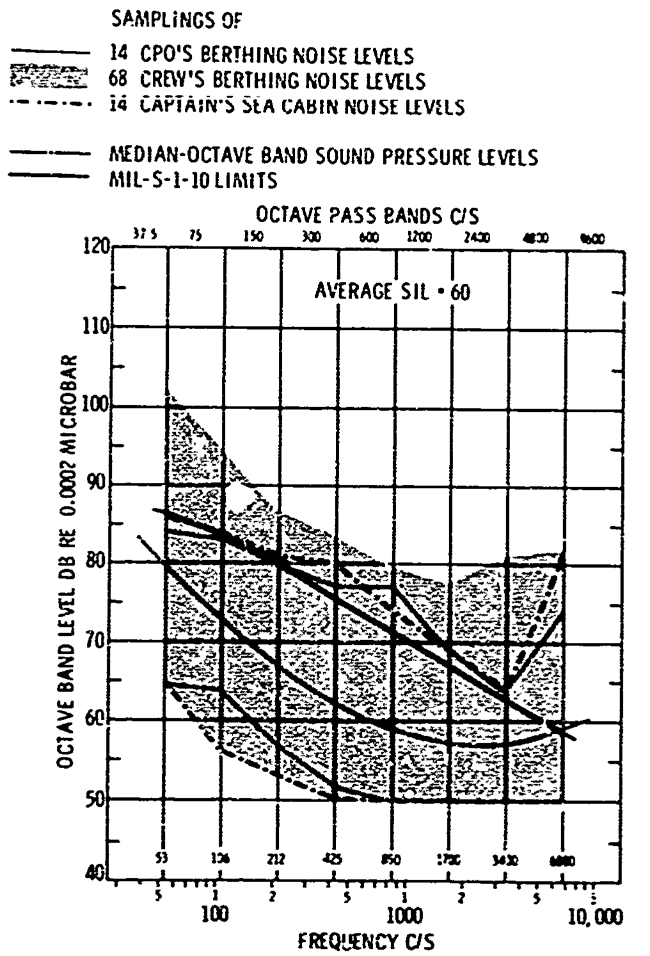


Figure II-6. The range for three Category B spaces (destroyers).

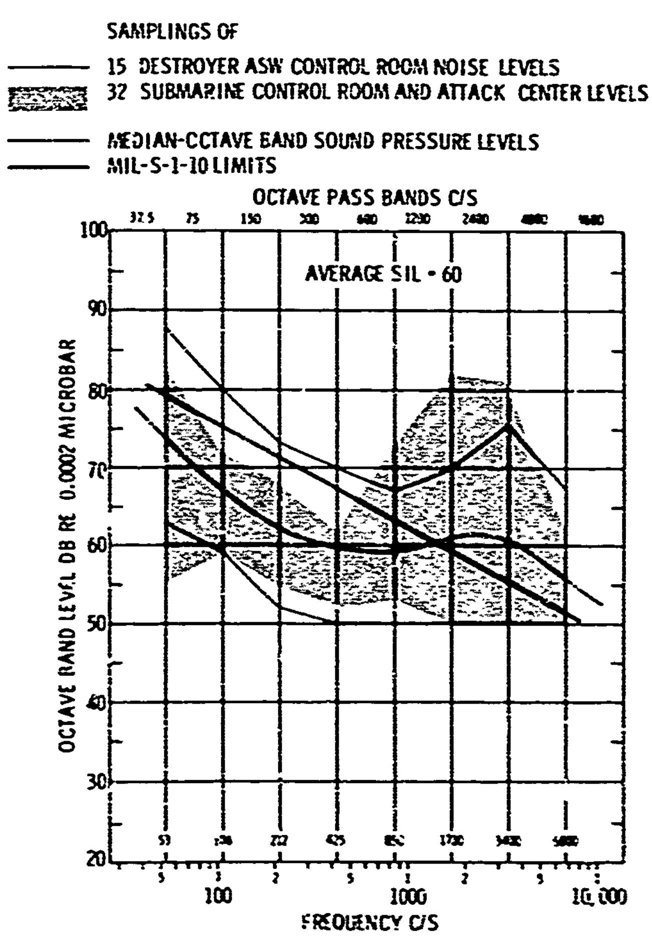


Figure II-7. The range for two Category C spaces.

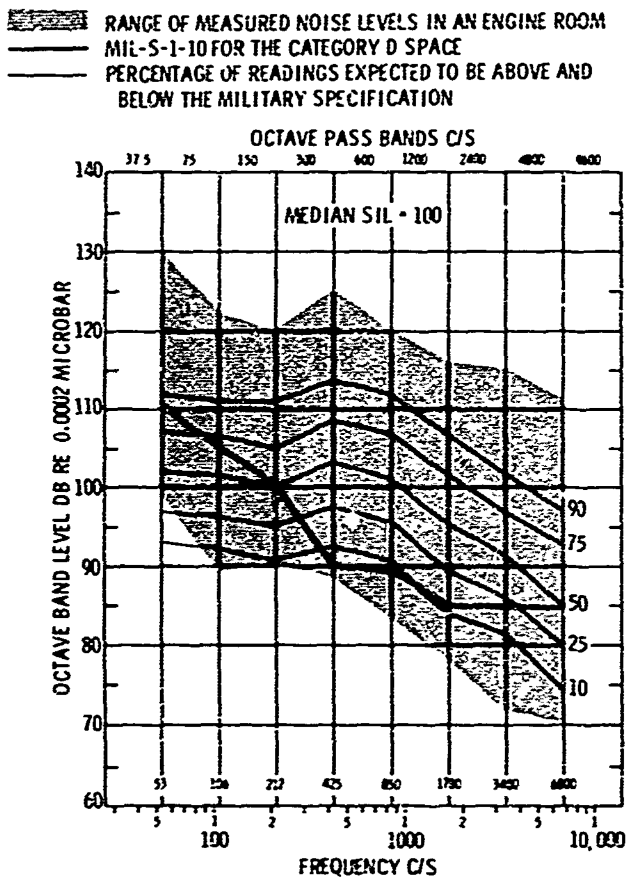


Figure II-8. The range for a Category D space (conventional submarine - full power run).

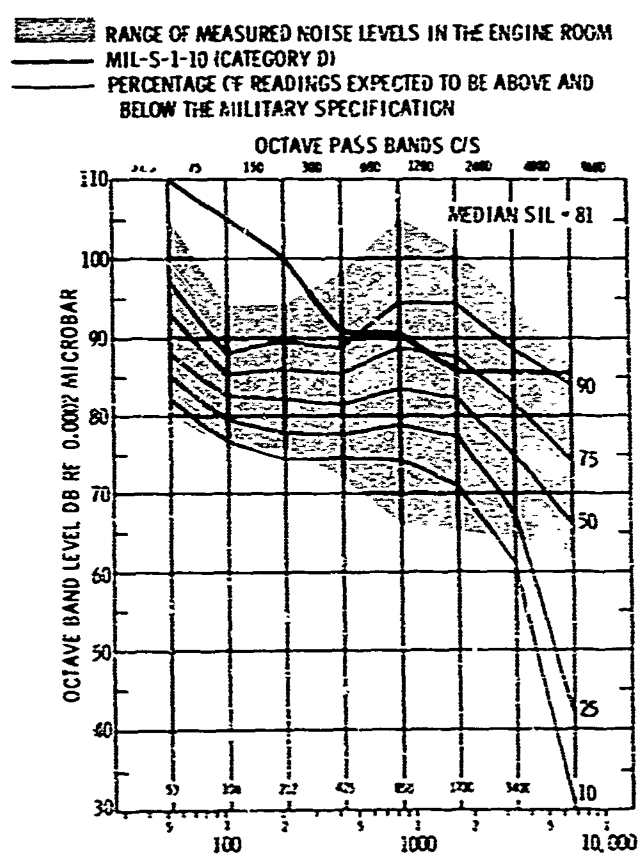


Figure II-9. The range for a Category D space (nuclear submarine - full power run).

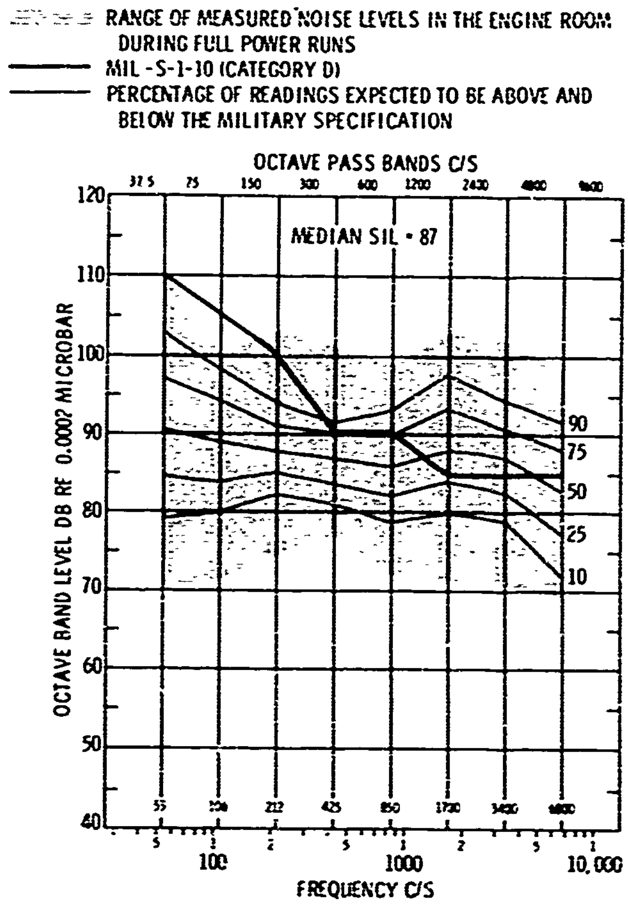


Figure II-10. The range for a Category D space (19 destroyers).

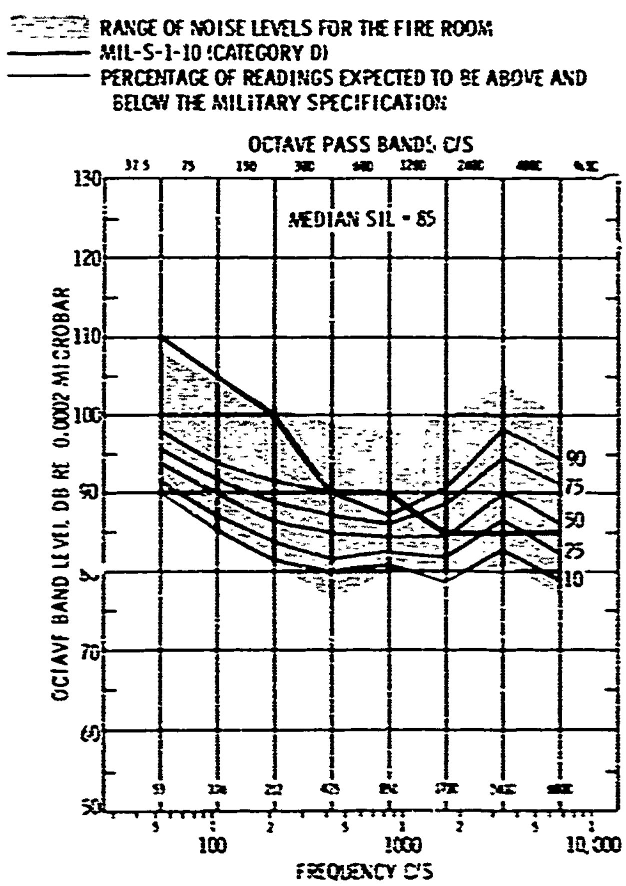


Figure II-11. The range for a Category D space (summary).

In summary:

1. Overall levels on three aircraft carriers, a missile cruiser, and two destroyers ranged from 70 dB to over 120 dB; median sound pressure levels for five different groups of measurements were 86, 93, 84, 90, and 80 dB.

2. The A-weighting levels taken aboard two carriers and a cruiser measured from 54 to 116 dB with a median value of 76 dB. Merchant ship values ranged from 65 to 105 dB.

3. Speech Interference Levels for two carriers and one destroyer measured from 55 dB to 100 dB with median values of 66, 80, and 73 dB. Ship specifications Section SI-10[2] give maximum permissible SIL's of 50, 55, and 60 dB for various sized category A compartments (in which speech communication is important) and 72 dB for category E compartments (in which deafness avoidance is a consideration, but a certain amount of speech communication is necessary). Obviously, with median measured SIL's of 66, 80, and 73 dB and estimated* median SIL's of about 66 dB and 62 dB for the data of figure II-1, the noise level in Navy ships is high enough to produce speech interference problems. It is not surprising that several noise surveys conclude with statements to the effect that SI-10 maximums were exceeded.[9,10,11]

It should be noted that although noise is a problem in the Navy, not all ships or compartments can be classified as noisy. Brief informal observations aboard the aircraft carrier USS RANGER (CVA 61), in June and August 1962, indicated a number of locations that were relatively free of noise (about 50 dB on the C scale of a sound-survey meter). The RANGER is a relatively new ship (commissioned in 1957), and from the observed widespread use of sound-absorbing and sound-isolating material, it appears that noise may be a lessened problem in some late model ships.

---

* For ship noise the median A level appears to be about 10 dB below the median C level and the median four-band SIL about 10 dB below the A level (fig. II-1 and II-2). This generalization should not be applied to individual measurements, but probably holds true for the general class of ship noise because of the all-pervading influence of low frequency noise from blowers and propulsion machinery.

NEL | A CONTRIBUTION OF THE NAVY ELECTRONICS LABORATORY, SAN DIEGO 52, CALIFORNIA

Reprinted from THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, Vol. 35, No. 9, 1328–1338, September, 1963

Physical Measurements of Equally Speech-Interfering Navy Noises

R. G. KLUMPP AND J. C. WEBSTER

U. S. Navy Electronics Laboratory, San Diego, California
(Received 3 March 1963)

The sound level of each of 16 diverse noises was adjusted so that loudspeaker-reproduced rhyme words at a level of 78 dB(C) at one meter were reduced to 50% word intelligibility. Spectrum analyses (octave band and 20-cps band), physical measurements (C, B, A, DIN 3 sound-level meter weightings), calculations [Speech Interference Levels (SIL)], and peak fittings [to noise criteria (NC)-type curves] were made to find which method came closest to agreeing that the noises were equally speech-interfering.

SIL calculations that included the octave 300–600 cps predicted the speech-interference aspects of the 16 noises best. The A and DIN 3 weighting networks, the conventional (600–4800 cps) SIL, and use of a restricted region (350–2800 cps) of NC-type curves gave the next best prediction. The poorest predictors were the B or C frequency-weighting networks and use of the NCA, NC, or ISO contours over their total frequency range.

INTRODUCTION

COMPARED to most civilian work areas, the ambient noise level in ship compartments is high. Noise in ship spaces comes from a variety of sources such as blowers, pumps, generators, public address systems, radio receivers, and men. With the ship underway, noise in lower-level compartments is augmented by noise from the propulsion gear and from hull vibration. Noise control is an ever-present problem on ships. Noise control on future ships is apt to present even greater problems due to the increased noise potential of future weapon systems. Simplified noise-rating methods will be needed.

This report re-examines the speech interference aspects of noise, with a view towards simplifying noise level measurement procedures. To accomplish this task: (1) noise levels in a number of shipboard spaces were measured aboard operating ships, (2) representative aircraft, ship, machinery, and office noises were collected and spectrum analyzed, (3) intelligibility tests were conducted to find the levels at which 16 of these noises interfered equally with speech, and (4) the 16 noises were subjected to a number of analytical physical tests and measurements to find a physical measurement that agreed that the noises were equally speech-interfering. Closely related studies were conducted to determine the relationship between noise level and speech level required for successful communication in a face-to-face situation,[1,2] to determine the variability among estimations of the average noise level obtained from observers reading moving-coil meters,[3] and to find psychophysical measurements of the noises that correlate with speech interference.[4]

I. COLLECTION OF NOISE SAMPLES

The first phase of the study to determine the speech-interfering properties of ship noises was to obtain a representative collection of recordings of such noises. The noise found aboard an aircraft carrier includes many of the noises found on other surface ships and in addition includes some peculiar to aircraft. For this reason, most of the ship noise samples were collected on aircraft carriers. Recordings were made in living spaces, communication spaces, engineering spaces, and on or

[1] J. C. Webster and R. G. Klumpp, "Effects of Ambient Noise and Nearby Talkers on a Face-to-Face Communication Task," J. Acoust. Soc. Am. 34, 936 (1962).

[2] R. G. Klumpp and K. H. Klumpp, "Using the Speech Level of a Talker to Assess the Speech Interference of Ambient Noise" (to be published).

[3] R. G. Klumpp and J. L. Leonard, "Observer Variability in Reading Noise Levels with Meters," Sound—Its Uses and Control 2, No. 4, 25–29 (1963).

[4] R. G. Klumpp and J. C. Webster, "Predicting Speech Interference from Physical and Psychophysical Measures of Ambient Noise," J. Acoust. Soc. Am. 35, 1116 (A) (1963).

about the flight deck with jet and propeller aircraft operating. For a ship's diesel engine noise a sample furnished by Lübcke[5] was used. Other recordings were made in machine shops and office spaces back in the laboratory. To complete the sampling, three standard laboratory noises were added: random noise, flat or unweighted; random noise shaped at +6 dB per octave; and random noise shaped at −6 dB per octave. These are later identified as TN Flat, TN+6, and TN−6.

Equipment used to record noise samples included a selected, omnidirectional Altec Lansing type 633A mi-

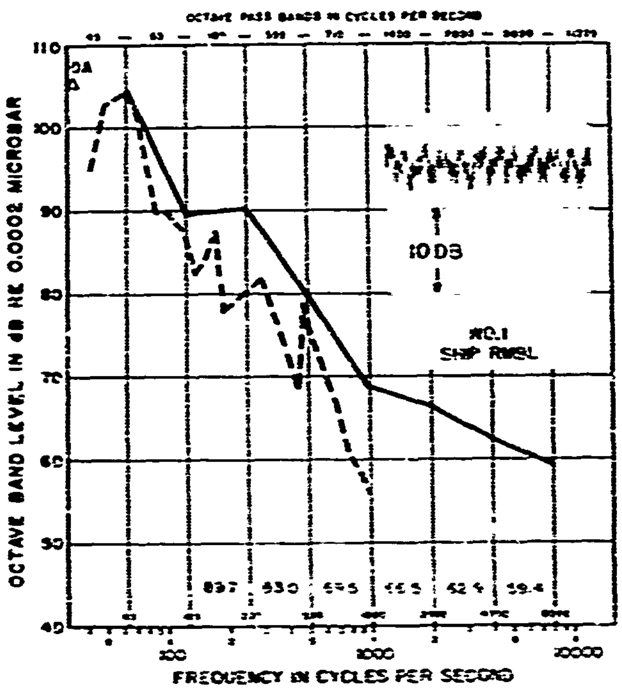


FIG. 1. Octave band and 20-cycle band levels of noise sample 1 (ship's rumble). This is a fairly steady, very-low-frequency rumble. A 15-sec log amplitude vs time trace is shown at the upper right. The over-all "C" level is shown at the extreme left. At the bottom are numbers representing the octave band levels in the octaves 150-300, 300-600, 600-1200, 1200-2400, 2400-4800, and 4800-9600 cps, i.e., those centered on the series 212, 425, etc. The octave band and 20-cycle band data are not directly comparable since these are acoustic analyses made on two different playback systems.

crophone, a General Radio Company type 1551A sound-level meter, and an Ampex Corporation type 601 recorder. In recording noise samples, the microphone was positioned at places normally occupied by people performing their jobs.

Of 150 recordings, 16 samples were selected to represent both intermittent and steady-state noises, to include both high- and low-frequency noises, and to include a number of noises commonly found aboard operating ships. A short description, together with log amplitude vs time trace and a spectral analysis, is shown for each noise in Figs. 1 through 16.

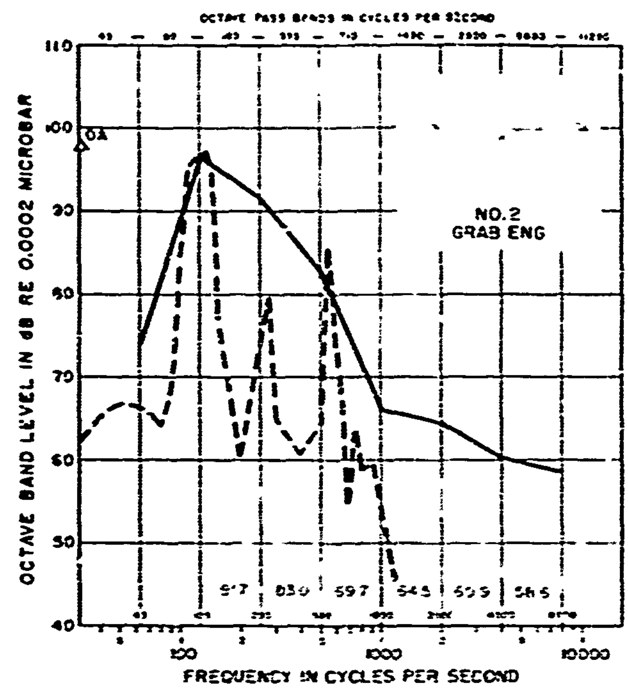


FIG. 2. Octave band and 20-cycle band levels of noise sample 2 (grab engine). This is an electrical motor of large size with strong low-frequency humlike components. See Fig. 1 for other details.

II. ANALYSIS OF THE 16 NOISES

The magnetic tape recordings of the noise samples were cut into loops of 10 to 15 sec duration and reproduced over an Ampex Corporation type 401 recorder. The signal from the recorder, amplified by two 20-W and one 30-W McIntosh Corporation amplifiers, was fed to an Acoustic Research loudspeaker, model 3,

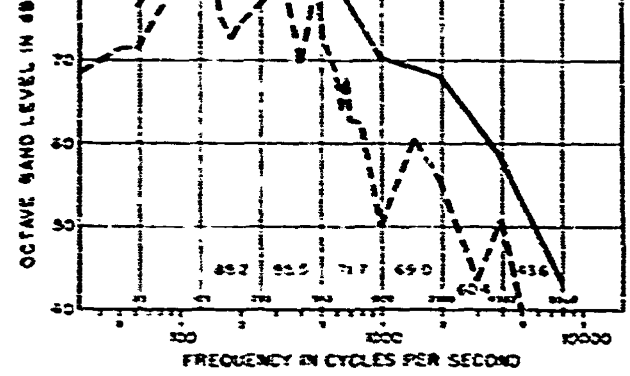


FIG. 3. Octave band and 20-cycle band levels of noise sample 3 (blower). This is a steady, low-frequency blower noise. See Fig. 1 for other details.

[5] Ernst Lübcke, "Tanker-Maschinenraum, 1958-106" (private communication).

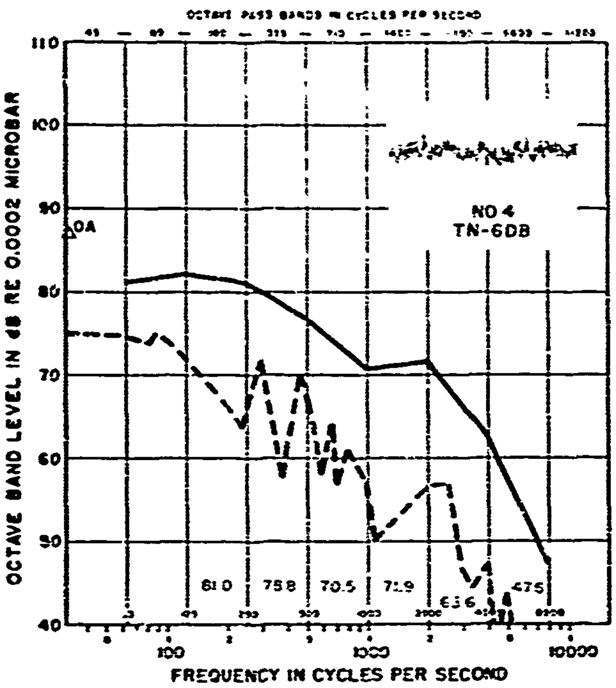


FIG. 4. Octave band and 20-cycle band levels of noise sample 4 (thermal noise sloped at minus 6 dB per octave). This is very steady, low-frequency noise. See Fig. 1 for other details.

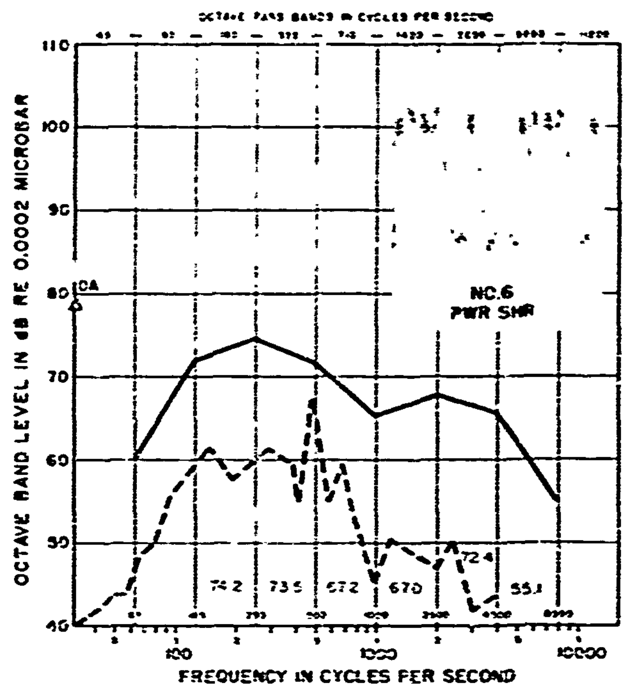


FIG. 6. Octave band and 20-cycle band levels of noise sample 6 (power shear), an extremely intermittent noise of a hydraulic power shear cutting metal, including clang of sheet metal dropping to floor. See Fig. 1 for other details.

and an Altec Lansing 604 loudspeaker system on one side and to an Altec Lansing 820 loudspeaker 18 feet away on the other side of a studio. The dimensions of the studio were 27 ft × 16 ft × 10 ft, and its reverberation time was about 0.3 sec at frequencies up to 1 kc/sec, rising to about 0.5 sec at 4 kc/sec. For analysis, the acoustic signal was picked up by a condenser microphone located midway between the two loudspeaker systems.

After the speech-interference tests were completed, octave band and other physical measurements of the acoustic reproduction of each of the 16 recorded samples were made. A Bruel and Kjaer type 2603 RMS meter and a pair of Allison filters with a combined rejection of

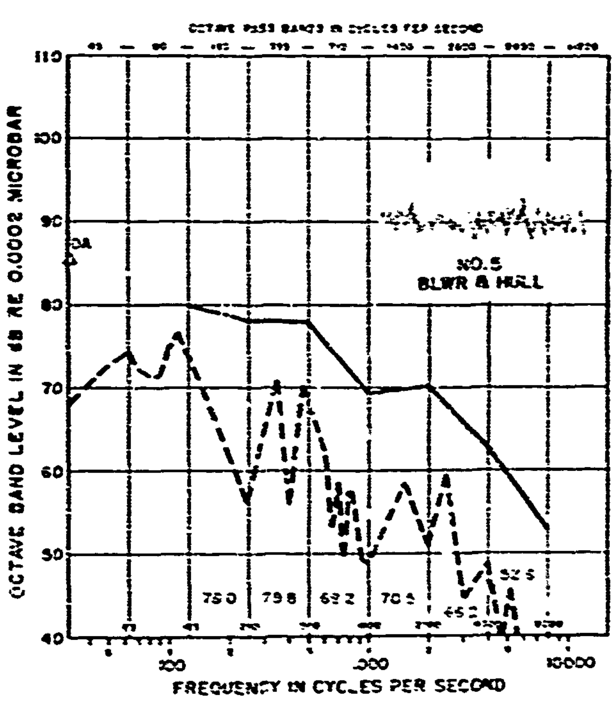


FIG. 5. Octave band and 20-cycle band levels of noise sample 5 (blower and hull), a fairly steady, low frequency rumble and machine noise. See Fig. 1 for other details.

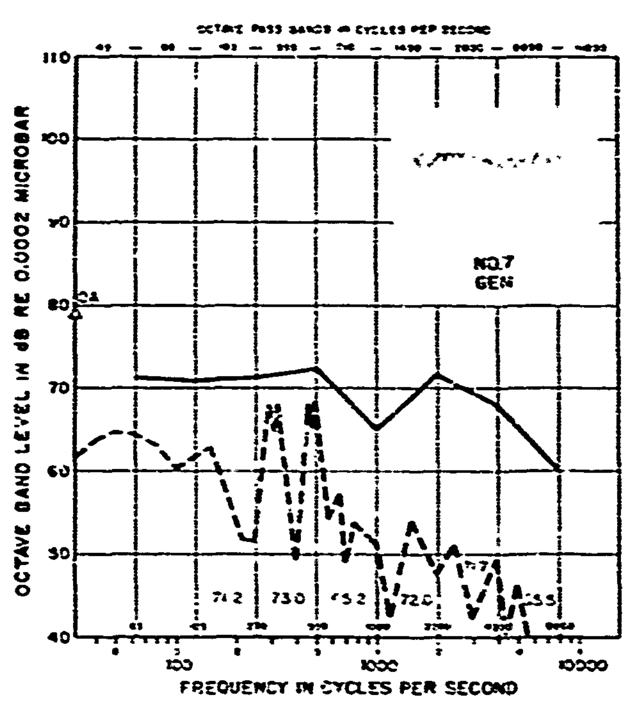


FIG. 7. Octave band and 20-cycle band levels of noise sample 7 (generator), which has whine and low-frequency components, with small regular changes in level. See Fig. 1 for other details.

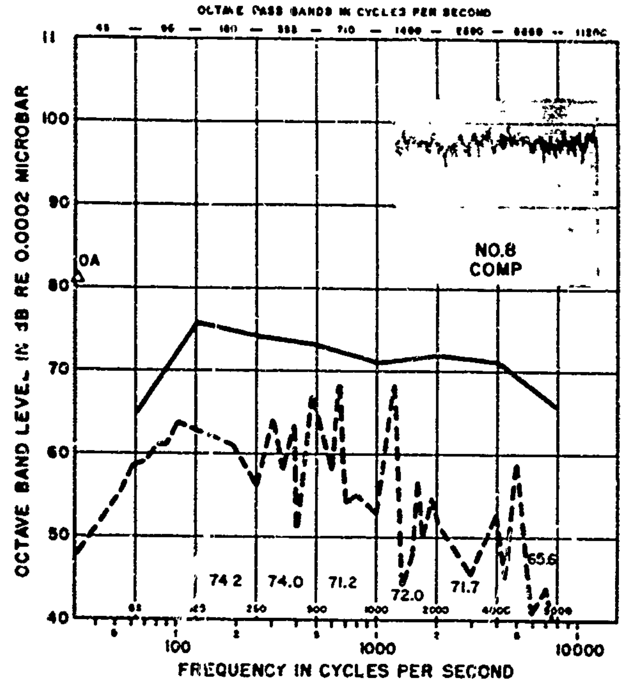


FIG. 8. Octave band and 20-cycle band levels of noise sample 8 (a compressor with rhythmic sound pattern). See Fig. 1 for other details.

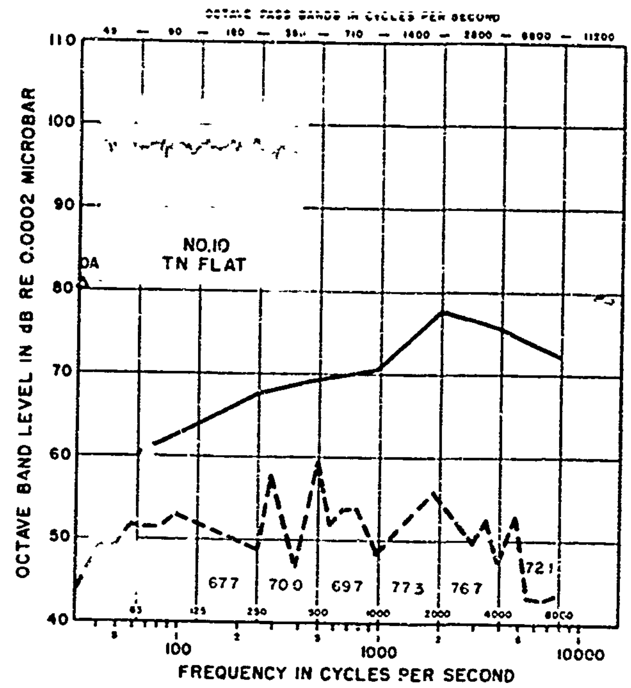


FIG. 10. Octave band and 20-cycle band levels of noise sample 10, a relatively flat, relatively constant level thermal noise with a wide-band hiss, relatively constant in level. See Fig. 1 for other details.

60 dB/octave were used to make these measurements. The octave band measurements are presented in Figs. 1 through 16, together with the log amplitude vs time traces. A 20-cps-band analysis, made before the final loudspeaker configuration was determined, is also given in these same figures for reference, although these data are not directly comparable to the octave band data.

The over-all levels (C weighting) on Figs. 1 through 16 represent equally speech-interfering levels; the determination of and the interpretation of these over-all levels are discussed later.

The noise samples were numbered on the basis of

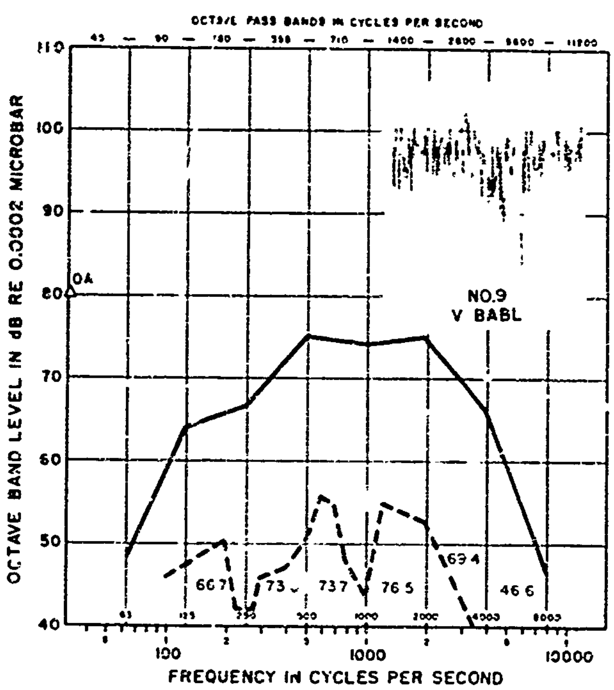


FIG. 9. Octave band and 20-cycle band levels of noise sample 9 (voice babble from Ref. 1). This consisted of pairs of communicators exchanging monosyllabic words, resulting in irregular level (20 dB) and frequency changes. See Fig. 1 for other details.

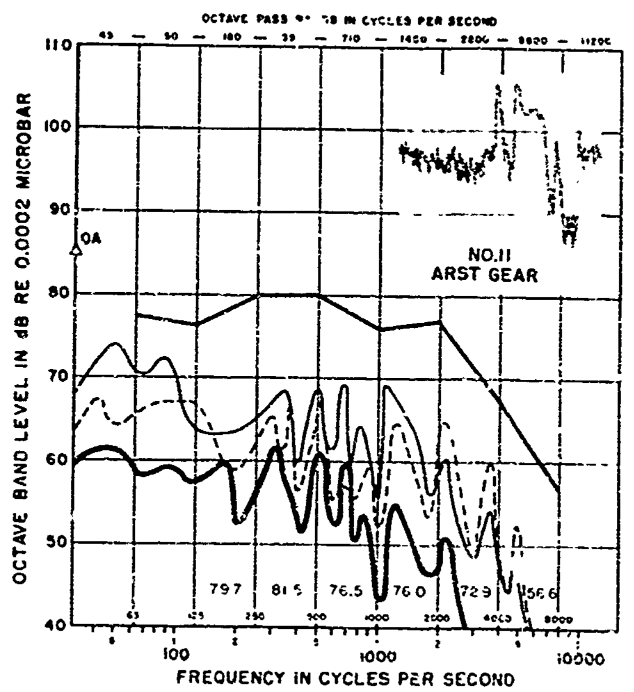


FIG. 11. Octave band and 20-cycle band level of noise sample 11 (arresting gear). This noise, which was recorded in an arresting-gear room of a carrier, has rumble, shriek, bang, and roar components, and fluctuates over a 20-dB range in an irregular fashion. See Fig. 1 for other details.

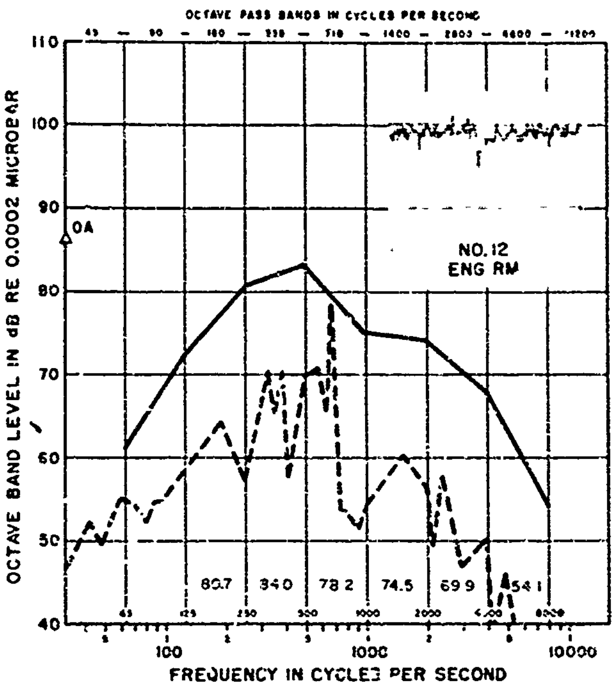


FIG. 12. Octave band and 20-cycle band levels of noise sample 12 (engine room). This complex machinery noise, furnished by Lübcke, is relatively steady in level. See Fig. 1 for other details.

frequency content, the low-numbered noises being rich in low-frequency sound and the higher-numbered noises rich in high-frequency sound. The noises numbered from about 6 through 12 have most of their energy in the mid frequency region. The amplitude vs time traces show that some of the samples are relatively constant in level while others vary as much as 20 dB in level

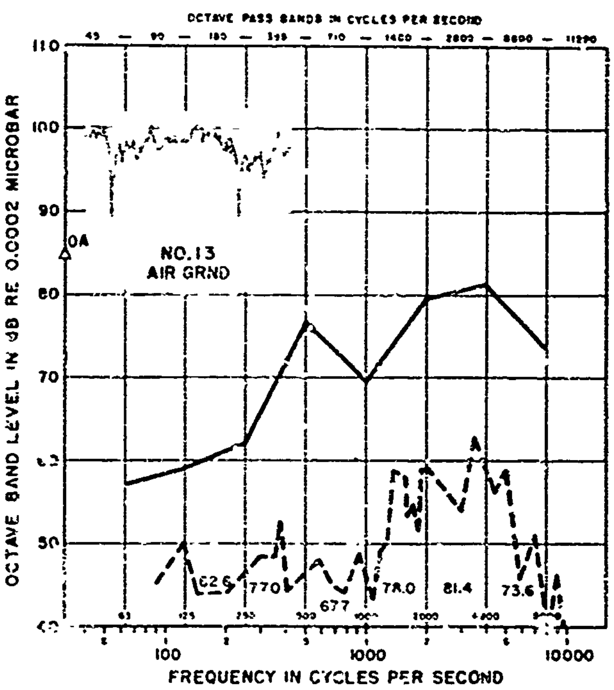


FIG. 13. Octave band and 20-cycle band levels of noise sample 13 (air grinder). This sound, similar to a "dental drill" sound, was produced by an air-driven grinder in contact with a metal sheet. See Fig. 1 for other details.

every 15 sec. This ranking system was not used as the order in which the speech-interference tests were run.

III. EQUALLY SPEECH-INTERFERING LEVELS OF NOISE

The aim of this portion of the study was to determine the speech-interfering effects of each of 16 noise samples by means of the procedure known as the speech-intelligibility test. In such a test, listeners are presented a mixture of speech and noise, asked to write down the speech they hear, and scored on the correctness of their responses. In this study, listeners heard constant level speech from a loudspeaker in the presence of each of 16

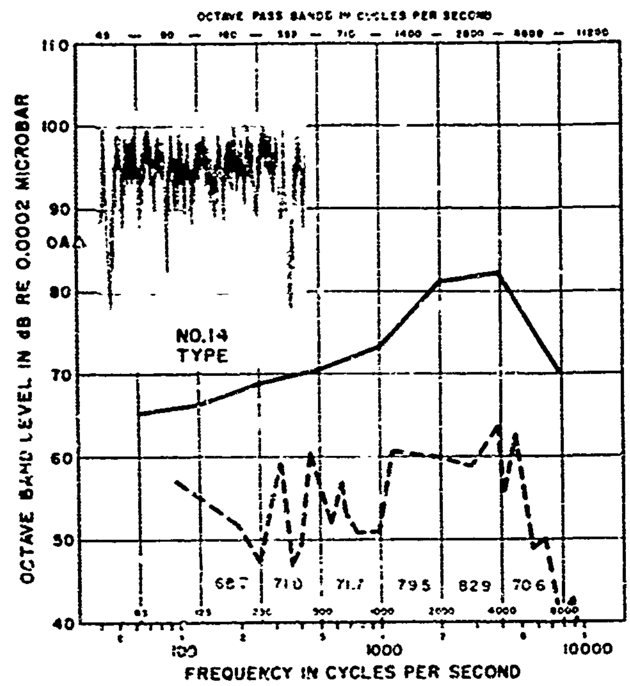


FIG. 14. Octave band and 20-cycle band levels of noise sample 14 (typewriter). This noise is strong in high-frequency components and, although the level fluctuates rapidly from moment to moment, the over-all level is fairly constant. See Fig. 1 for other details.

different ambient noises. The level of each of the noises was adjusted to give equal speech interference, i.e., 50% of the words heard correctly. The levels of the 16 noises were then compared to find the measurement procedure or system according to which the measured levels were most nearly equal.

Procedure

Test words were recorded using a Radio Corporation of America type BK6B lavalier microphone in order to minimize changes in recorded voice level resulting from movements of the talker with respect to the microphone. The single male talker used had been found, in previous noise-masked tests, to be average or slightly above average in intelligibility when compared with other laboratory personnel. To simulate the performance of a talker in noise, the talker spoke at a level approximately 8 dB above normal during recording. This increased vocal output was achieved by having the talker first deliver words with normal vocal effort without watching a VU meter in the recording system. The average VU meter deflection output was noted and 8 dB of attenuation was added to the meter circuit. The talker was then instructed to record using the VU meter and to adjust his vocal output to the level which produced the same meter deflection obtained in the normal voice condition. Actual levels were set on the carrier word "write," which was given before each test word.

Rhyme[6] words were recorded at a rate of one word every two seconds. This two-second rate is faster than that used by Fairbanks but has been found to be satisfactory in previous tests.[7] The rhyme test is essentially a test of consonant discrimination with the listener required to fill in the first letter of an incomplete monosyllabic word.

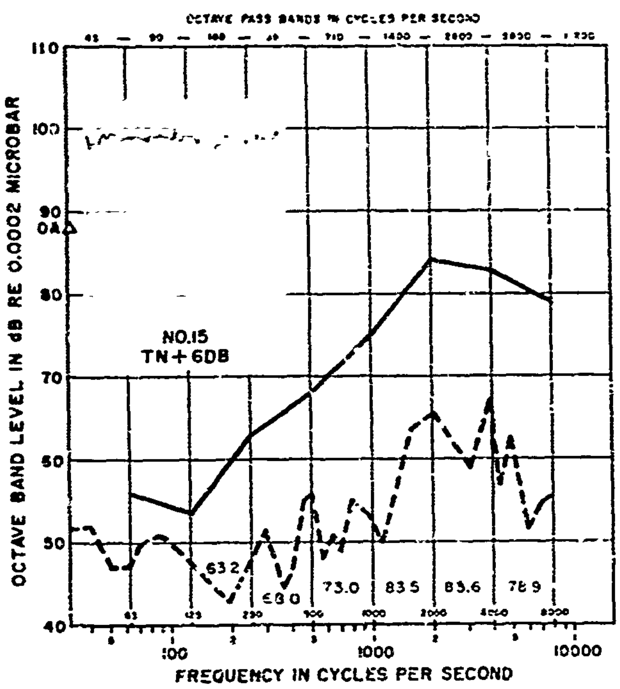


FIG. 15. Octave band and 20-cycle band levels of noise sample 15 (thermal noise sloped at plus 6 dB per octave). This is a high-frequency hiss with a relatively constant level. See Fig. 1 for other details.

One hundred lists of 50 rhyme words each were recorded over a 4-day period. After elimination of all lists with errors or ambiguous pronunciations, five blocks of 10 lists each were assembled, with each block having equal representation from each of the four days of recording.

For the intelligibility tests the lists were played back

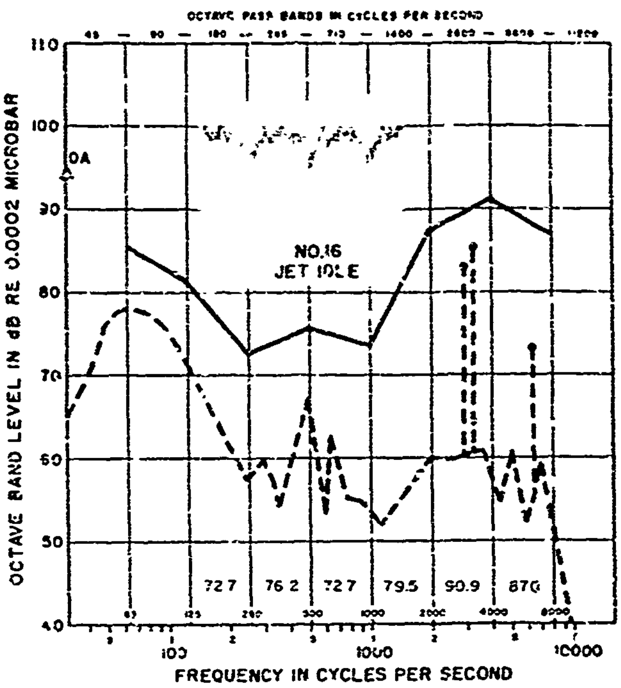


FIG. 16. Octave band and 20-cycle band levels of noise sample 16 (jet idle). This recording of a jet engine on the flight deck of a carrier has strong whine components and has several abrupt drops in level.

over a University MLC loudspeaker mounted on an adjustable stand. The frequency response of this loudspeaker, measured 3 ft from the face of the speaker (head position of the listener), is shown in Fig. 17. Playback level of the speech at the listener's position as measured on a sound-level meter set to "C" scale, "fast," was approximately 78 dB.

Ambient noise signals were provided by the playback system described earlier in Sec. II, and the physical measurements of the noises (Figs. 1–16) were made with the microphone at a position corresponding to the

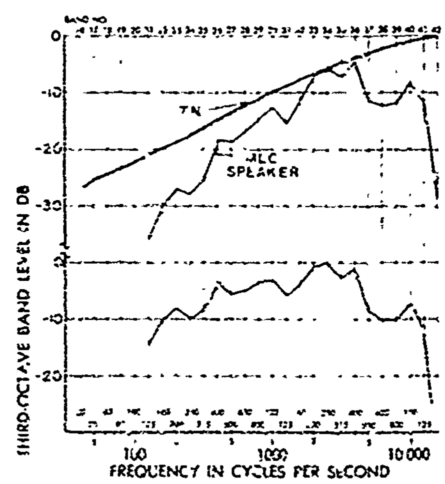


FIG. 17. Frequency response, in third-octave bands, of loudspeaker used to reproduce rhyme words. (Add 4.9 dB to obtain octave-band level.) *Upper curve:* Electrical response of thermal noise. *MLC curve:* Acoustical output of MLC speaker to thermal-noise input. *Bottom curve:* Difference between electrical input and acoustical output.

[6] G. Fairbanks, "Test of Phonemic Differentiation; The Rhyme Test," J. Acoust. Soc. Am. 30, 596 (1958).

[7] J. C. Webster and R. G. Klumpp, "USNEL Flight Deck Communications System. Part 2. Noise and Acoustic Aspects," U. S. Navy Electronics Laboratory Report 923, AD-250 286 (1960), footnote, p. 19.

center of the listener's head. The noise measurements, therefore, represent some sort of an "average" noise at the listener's head position.

The listener was seated at right angles to and midway between the two speaker arrays located 18 ft apart. Although the noise sources were located to each side of the listener, noise was localized as coming from above (equal, in-phase signals at each ear). The loudspeaker emitting the word lists was located in front of the listener, at ear-level height, and with the face of the speaker approximately 36 in. from the ears of the listener. The listener's chair was fixed in position and, because of the proximity of the small table on which he wrote his answers, he could not move more than an inch or two in any direction.

The intelligibility tests in this experiment differ from the usual earphone or single loudspeaker tests in that the speech and noise sources were spatially separated. This means that the listener could "selectively attend" to the speech. Precisely how this might affect the results is not certain. The listener could move his head over a restricted area and hence could position his head with respect to the speech source in front of him and the noise sources at both sides. In this manner he could maximize speech-to-noise ratio. In particular, this procedure might be effective against high-frequency-noise, standing-wave patterns, such as existed in the jet idle noise (#16), which contained several strong, high-frequency tonal components.

All listeners were college students with normal hearing (no loss greater than 10 dB in either ear from 125 to 6000 cps as measured on a Békésy-type audiometer). Five males and three females in the 17–21-year age bracket served.

The testing sequence for each listener was as follows: (1) audiometric test, (2) practice in quiet (10 word lists), (3) practice in noise (10 lists in each of 4 levels of unweighted random noise), (4) main tests in which 10 lists were presented in each of the 16 noises, and (5) a retest with the same block of 10 lists under the same conditions as in the latter portion of the "practice in noise" sequence to assess the effects of learning.

Listeners were tested individually. During the main tests, a single session for a given listener consisted of the presentation of 10 lists in a noise, a 15-minute rest, and then the presentation of 10 more lists in a second noise. Each listener served in two sessions per day, one in the morning and one in the afternoon, until the required 12 test sessions were completed.

Blocks of 10 lists were counterbalanced among listeners and noises, and the sequence of noise samples was counterbalanced among listeners and afternoon and morning sessions.

Results

Table I lists percentage of words correct obtained from each of the 16 noises. Each percentage is based on 4000 words (8 listeners×10 lists×50 words per list). Column 4 lists the sound-pressure level of each noise (C scale of sound-level meter) as it was set during the intelligibility tests. Based on the assumption that a change of 1 dB in speech-to-noise ratio would produce a change of 5% in the percentage of words heard correctly, column 5 lists the amount in dB to be added to column 4 to produce a score of 50%.[3] The sum of columns 4 and 5, column 6, yields equally speech-interfering levels of these noises expressed in terms of the C weighting network of the sound-level meter.

TABLE I. Percent words correct vs over-all noise level.

	Noise	Percent Correct	C Scale reading in dB	Correction in dB	Equally Interfering "C" Level in dB
1	Ship Rumble	40.5	107.2	−1.9	105.3
2	Grab Engine	52.5	97.1	0.5	97.6
3	Blower	50.1	92.3	0.0	92.3
4	TN-6	42.6	88.3	−1.5	86.8
5	Blower & Hull	52.2	84.9	0.4	85.3
6	Shear	54.5	77.9	0.9	78.8
7	Generator	34.7	82.3	−3.0	79.3
8	Compressor	49.3	81.2	0.2	81.0
9	Babble	51.9	79.9	0.4	80.3
10	TN Flat	53.1	80.1	0.6	80.7
11	Arresting Gear	44.8	86.3	−1.0	85.3
12	Engine Room	55.7	85.1	1.2	86.3
13	Air Grinder	48.0	84.7	−0.4	84.3
14	Typewriter	52.7	85.9	0.5	86.4
15	TN+6	52.8	87.7	0.6	88.3
16	Jet Idle	51.4	93.5	0.3	93.8
	Average	49.2			

An average score of 99.8% was obtained for the 500 words presented to each listener in the quiet (6 listeners heard every word correctly, one person missed 1 word in 500, and one missed 8 words in 500).

Pre- and post-test scores for the group of 8 listeners were 65.7% and 66.8%, respectively. The difference 1.1% was not statistically significant. A difference equal to or greater than 4.7% would be required to be significant at the 5% level of confidence.

An analysis of variance was performed on the data and it was found that differences among subjects, between scores obtained in the morning and scores obtained in the afternoon, between scores obtained for a noise when it was presented first in a session and scores

[3] (a) C. V. Hudgins, J. E. Hawkins, J. E. Karlin, and S. S. Stevens, "The Development of Recorded Auditory Tests for Measuring Hearing Loss for Speech," Laryngoscope 57, 57 (1947). They show a 4% per dB change in monosyllabic scores in the vicinity of 50%. Fairbanks (Ref. 6) states the slope to be 3% per dB over an extended (33% to 84%) range, but his data when restricted closer to the neighborhood of 50% show more like a 5% per dB slope; (b) W. E. Montague, "A Comparison of Five Intelligibility Tests for Voice Communication System," NEL Report 977, PB157 229, AD254-545 (1960). He used the rhyme words at the rate used in this study, and gets a slope of about 7% per dB. Had either 7% or 4% been used instead of 5%, the difference in correction would have been less than 1 dB for the noise deviating most from 50%: namely, noise #7, generator.

obtained for a noise when it was presented second in a session, and among the variability of intelligibility scores obtained with each of the 16 noises were not statistically significant.

IV. PHYSICAL PROPERTIES OF EQUALLY SPEECH-INTERFERING NOISES

It is apparent from column 6 of Table I that the "C" level of a sound-level meter does not predict the speech-interfering properties of noise. These noises were adjusted in level to give equal scores on a speech-intelligibility test, but the levels as measured vary from 78.8 to 105.3 dB(C) with a standard deviation of 6.9 dB.

It is neither new nor surprising to find that the dB(C) level of a noise does not predict its speech-interference properties. Many schemes have been worked out that take into account the two prime determiners of speech intelligibility in noise: namely, the spectrum of the noise and the speech, and the difference in level between the speech and the noise. The most sophisticated of these schemes, the articulation index (AI), is not discussed in this paper, but is discussed in the following paper.[9] Here, measurements and calculations are confined to simple measures on the noises themselves: the noises that have been equated in level to be equally speech-interfering.

Three methods of measurement are discussed: the reading of a sound-level meter with weighting networks, the fitting of plotted noise spectra to families of noise criteria (NC) curves, and the calculation of speech interference level (SIL). Good summary statements of SIL with the limitations and assumptions detailed are given by Rosenblith and Stevens[10] and Beranek.[11] The general idea is that if the level of noise in each of three,[12] four,[13] or five[14] octave bands is known speech levels needed for adequate communication at specified distances can be predicted. Noises with different spectra and levels that yield the same SIL would be equally speech-interfering. That is, the level of speech needed to be understood in a given SIL would be the same regardless of the combination of spectra and levels that produced the SIL.

For an office space to be acceptable it must be quiet enough to allow easy face-to-face or telephone voice communications. A set of noise criteria (NC) curves which would predict office acceptibility would indirectly predict adequate voice communications. Beranek[15] has developed such a set of NC curves having numbers assigned which are the average of levels in the 600–1200-, 1200–2400-, and 2400–4800-cps octave bands. Any noise that does not exceed a level in any octave band greater than a specified NC curve level is assigned an SIL number corresponding to that NC curve. If the spectrum of a noise is known, its SIL can be estimated from the NC curves. This fitting of a noise spectrum to a set of noise criteria contours will be called the "tangent-to-curve" method, and constitutes a second way of estimating the speech-interfering property of a noise from knowledge of its physical properties.

Young[16] has pointed out "... that the NC curves are similar in shape to the inverse of the A-weighting." The manner of use of the NC curve makes it comparable to frequency weighting having a shape inverse to the NC; therefore A levels should correlate closely with NC, at least for sounds having tangency to the NC curve over a limited frequency band. The latter qualification is required because the sound-level meter integrates power over the entire spectrum after weighting, whereas the tangent-to-curve technique responds only to the maximum tangency level, whether it be a single point or parallels the entire NC curve. Considering the response of the ear to masking and loudness, one would expect the integrating measure to be superior to the non-integrating one.

Since all the 16 noises were adjusted in level to produce equal speech intelligibility in part III of this experiment, the effectiveness of any of the prediction methods (SIL, NC, A sound level, etc.) can be assessed by its ability to yield an equal number on each of the 16 noises. The standard deviation around the mean of the 16 measures on SIL, NC, A, etc., is thus an inverse measure of which scheme actually predicts speech intelligibility the best.

Table II shows the various measures associated with the 16 equated noises. Columns 1 and 2 identify the noises. Column 3, headed "weight," lists an estimate made by three individuals, experienced in noise measurements, of the relative frequency of occurrence of each type of noise. That is, in a general sampling of ship noises, noises 1, 3, and 5 would compose about a third of all noise samples (11/33); noises 7, 8, 9, and 14 another third; and the remaining nine noises would be typical of the remaining third of noise types. Columns 4, 5, 6, and 7 show the over-all levels as measured on the C, B, A, and DIN 3 (Ref. 17) weighting networks of a sound-level meter.

[9] J. C. Webster and R. G. Klumpp, "Articulation Index and Average Curve-Fitting Methods of Predicting Speech Interference," J. Acoust. Soc. Am. 35, 1338 (1963).

[10] W. A. Rosenblith and K. N. Stevens, *Handbook of Acoustic Noise Control. Vol. 2: Noise and Man.* Wright Air Development Center Tech. Rept. 52-204 (1953).

[11] L. L. Beranek, *Acoustics* (McGraw-Hill Book Company, Inc., New York, 1954), pp. 419–420.

[12] L. L. Beranek, "Airplane Quieting II—Specification of Acceptable Noise Levels," Trans. Am. Soc. Mech. Eng. 69, 97 (1947).

[13] M. Strasberg, "Criteria for Setting Airborne Noise Level Limits in Shipboard Spaces," Bureau of Ships Report No. 371-N-12 (1962).

[14] J. M. Pickett and K. D. Kryter, "Prediction of Speech Intelligibility in Noise." Air Force Cambridge Research Center, Tech. Rept. 55-4 (1955).

[15] L. L. Beranek, "Revised Criteria for Noise in Buildings," Noise Control, 3, No. 1, 19–27 (1957).

[16] R. W. Young, "Don't Forget the Simple Sound-Level Meter," Noise Control, 4, No. 3, 42–4? (1958).

[17] DIN 5045, April 1962, see A. Peterson and P. V. Bruel, "Instruments for Noise Measurements," in *Handbook of Noise Control*, edited by C. M. Harris (McGraw-Hill Book Company, Inc., New York, 1957), Chap. 16, p. 16–13, Fig. 16.12(b).

TABLE II. Physical measures or calculations in dB.

Noise	Weight	Frequency weighting C	B	A	DIN 3	NC	Tangent-to-curve NCA	ISO	ISO(R)	SIL calculations 600–4800 cps	300–4800 cps	300–2400 cps	3 band 500–2000 cps	4 band 250–2000 cps
1 Ship Rumble	3	105.3	96.5	86.3	89.3	91	80	88	77	66.3	70.5	73.0	71.6	76.3
2 Grab Engine	1	97.6	94.3	86.3	80.5	85	81	87	89	64.4	69.0	71.7	71.1	76.3
3 Blower	2	92.3	90.3	85.1	81.1	83	79	83	82	67.0	71.7	75.4	73.7	77.3
4 TN 6	2	85.8	83.5	79.3	77.3	75	73	74	74	69.3	71.7	73.7	73.1	75.1
5 Blower & Hull	5	85.3	82.3	78.8	77.1	75	73	75	75	68.6	71.4	73.2	72.3	73.7
6 Gear	1	78.8	78.6	75.8	74.3	69	69	71	71	68.9	70.0	69.2	68.5	70.0
7 Generator	2	79.3	79.3	78.6	77.9	72	72	74	74	69.9	70.7	70.1	69.6	70.0
8 Compressor	3	81.0	80.6	78.5	78.1	72	73	75	74	71.4	72.1	72.4	72.0	72.6
9 Babble	3	80.3	81.3	80.8	80.3	76	76	77	77	73.2	73.2	74.4	74.8	72.8
10 TN Flat	1	80.7	82.3	81.6	81.6	78	78	80	80	74.6	73.4	72.3	72.6	71.4
11 Arresting Gear	1	88.3	84.3	83.8	82.8	78	78	79	79	75.2	76.8	78.1	77.3	77.9
12 Engine Room	2	86.2	86.3	84.3	82.8	80	78	81	81	74.2	76.7	78.9	77.4	78.2
13 Air Grinder	1	84.7	85.8	84.8	85.3	83	84	85	82	75.7	76.0	74.2	75.3	[illegible]
14 Typewriter	3	86.4	87.9	87.4	87.9	85	85	87	83	78.0	76.3	74.1	74.8	73.3
15 TN +6	1	88.3	90.1	89.8	89.8	85	85	87	87	80.0	77.0	74.8	75.9	72.7
16 Jet	1	93.8	95.1	94.3	94.3	93	94	95	90	81.0	79.8	76.1	79.0	77.4
Range		26.5	18.2	18.5	20.0	24.0	25.0	24.0	19.0	16.6	10.8	9.7	10.5	7.9
Mean (16)		87.0	86.2	85.5	82.0	80.2	78.7	81.1	79.1	72.4	73.0	73.9	73.7	74.2
Stan. Dev. (16)		7.4	5.5	4.7	5.0	6.8	5.2	6.4	4.8	4.8	3.1	2.5	2.8	2.7
Stan. Dev. (33)		6.7	5.4	4.2	4.3	6.5	4.7	5.9	4.3	4.3	2.6	2.1	2.4	2.4
Mean (33)		87.2	85.9	82.7	81.0	79.5	77.6	80.3	78.2	71.4	73.0	73.8	73.4	74.2
Rank Order 16		13	10	5	8	11	9	12	6.5	6.5	4	1	3	2
Rank Order 33		13	10	5	7	11	9	12	7	7	4	1	2.5	2.5
Group Rank		3		2		3			2		1			

Columns 8 and 9 result from overlaying the noise spectra (Figs. 1 to 16) onto families of NC and NCA (noise criteria alternate) contours. Since there are no NC or NCA contours above 70, ratings are extrapolated on the assumption that contours above 70 would be drawn parallel to the 70 contours.

For columns 10 and 11, the noise spectra were overlaid onto a whole family of ISO contours.[15] The contour (usually interpolated) just tangent to the spectral peaks was noted in column 10. In column 11 are the results when only the range between 500 and 2000 cps was used; this is the restricted range ISO application and is labeled the ISO(R) method.

Column 12 shows the simple arithmetic average of the sound-pressure levels (measured in dB) in the octaves 600–1200, 1200–2400, and 2400–4800 cps. This is the common SIL calculation.[19] Column 13 shows the Strasberg[11] SIL calculation which is the arithmetic average of the dB levels in the octaves 300–600, 600–1200, 1200–2400, and 2400–4800 cps. Column 14 is an average of the sound-pressure levels in three octaves from 300 to 2400 cps.

Columns 15 and 16 are averages of sound-pressure levels in octaves centered at the "preferred frequencies"[19] of 500, 1000, and 2000 cps (col. 15, 3 band), or 250, 500, 1000, and 2000 cps (col. 16, 4 band).

Below each column is listed the range of measurements (the difference between the largest and smallest number), the mean measure, and the standard deviation of the measurements. Two means and standard deviations are shown, one based on the 16 measures as listed and a second based on the weights (which total 33) given in column 3. That is, the mean (33) and standard deviation (33) are based on 3 noises like #1, one like #2, 3 like #3, etc. The ideal calculation scheme or measurement system would show a zero for both range and standard deviation. The most consistent system is the one with the smallest standard deviation.

The measurement or calculation schemes can be arranged in rank order of excellence by assigning the rank 1 to the smallest standard-deviation measure and 13 to the highest. This is done on Table II at the bottom, but the ranks must be interpreted with caution. Each standard-deviation measure has its own error of measurement [the standard error of a standard deviation is $(2N)^{-\frac{1}{2}}$ times the standard deviation]. If this is taken into account, the measurement schemes can be rank-ordered into more meaningful groups as follows: (1) all SIL's except 600–4800 cps; (2) SIL (600–4800 cps), ISO(R), A, DIN 3; (3a) B, NCA, and (3b) ISO, NC, and C.

From this grouping it is apparent that all SIL methods, except the conventional 600–4800-cps one, predict the speech-interfering properties of noise best. The A- or DIN 3-weighted sound-level measurements, the conventional (600–4800 cps) SIL, and the ISO(R) curve are next best, and the remaining tangent-to-curve methods and the B- and C-weighted sound levels are the worst.

These results can be better visualized in Fig. 18, which is a plot of the data in Table II. In Fig. 18 it will be noted that the frequency weighting networks give higher readings than the SIL measurements. This is because (1) the whole spectrum is measured and (2) the levels combine as the squares of the sound pressures. In the SIL measurements, sound-pressure levels above

[15] International Standards Organization Technical Committee-43 (Secretariat-159)235, Draft Secretariat Proposal for Noise Rating Numbers with respect to Conservation of Hearing, Speech Communication and Annoyance (Aug. 1961).

[19] "American Standard for Preferred Frequencies for Acoustical Measurement," S1.6-1960 (American Standards Association, New York).

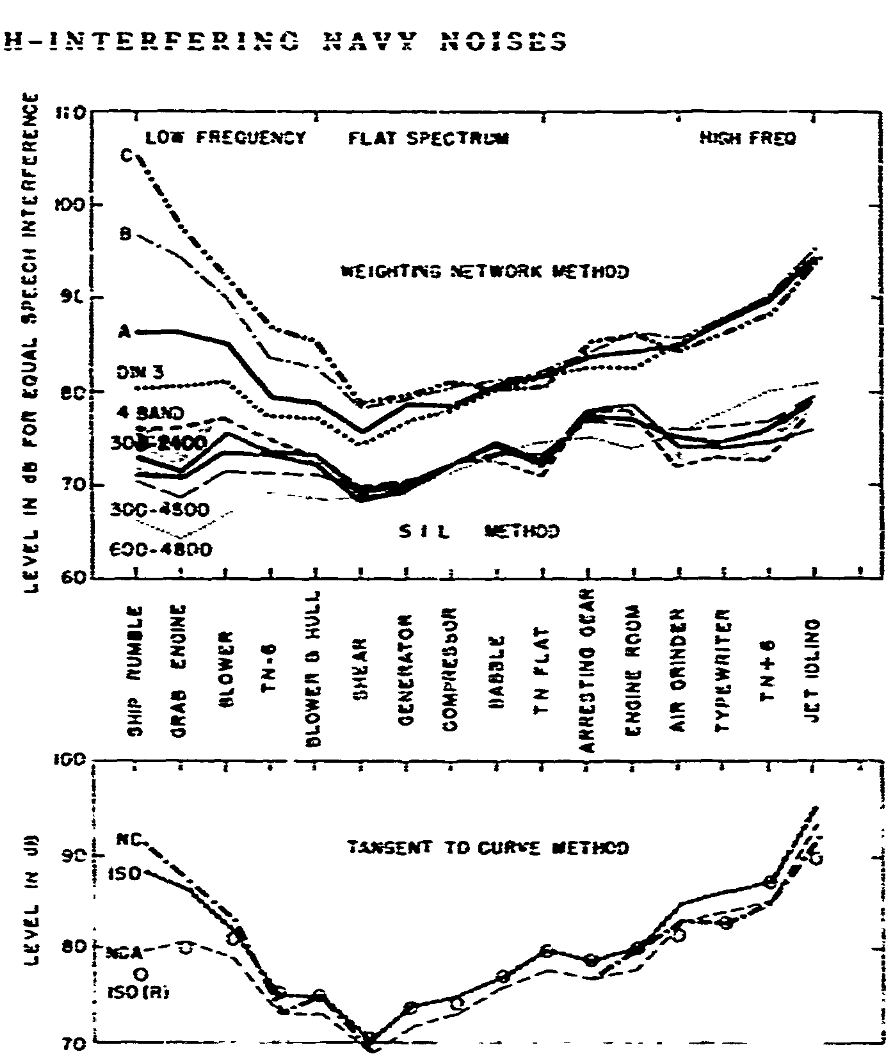


FIG. 18. Plot of data in Table II. Measured or calculated levels in dB for 16 noises equated in level to be equally speech-interfering.

and below a certain cutoff point are ignored and the octave band levels are arithmetically averaged: they are not summed.

Inspection of Fig. 18 shows that the noises group themselves into roughly four clusters. The first five noises are characterized by high C readings followed in order by B, A, and DIN 3. These are noises with considerable low-frequency energy. This is verified by noting that the SIL over the 300–2400-cps octaves exceeds the 300–4800-cps SIL, and both exceed the 600–4800-cps SIL.

The next group of five noises is characterized by nearly equivalent readings among the C, B, A, and DIN 3 network, and among the three SIL-type averages.

The arresting-gear and engine-room noises are malcontents. The relative levels among C, B, A, and DIN 3, and among the SIL's are like the low-frequency noises. But the magnitudes of these same measures lie midway between the flat spectrum noises and the remaining four high-frequency noises. One of these noises, the arresting gear, is very intermittent and is hard to specify and to measure. The other, Lübcke's engine room noise, is peaked in the mid-frequencies and is neither a high- nor low-frequency noise, nor a flat one.

The last four noises are characterized by approximately equal readings on networks A, B, and DIN 3, and all of them greater than C, and by the 300–2400-cps SIL being less than the 300–4800-cps SIL, and both less than the 600–4800-cps SIL. These are predominantly high-frequency noises.

Three basic methods of specifying the physical characteristics of these 16 equally speech-interfering noises have been shown in Table II and Fig. 20. These are (1) the frequency-weighting network method (A, B, C and DIN 3) as used on sound-level meters; (2) the SIL calculation methods [arithmetic average of sound-pressure levels (measured in dB) in contiguous octave bands]; and (3) the tangent-to-standard-curve method (adjust peaks of plotted spectra to sets of curves). Of these three methods, the SIL measurements show the lowest values of dispersion. The A- and DIN-3-weighting networks are the best of the meter network methods. The best tangent-to-curve methods are those utilizing a restricted range or a steeply rising low-frequency contour.

These three methods work in different ways and it is pertinent to point out how they differ. The simplest in concept, but the worst in predictive ability, is the tangent-to-curve method. In this method only the noise peak that first becomes tangent to the generalized curve(s) determines the measure. Any pure tone component, or any restricted band component, that differs drastically from its surround specifies this rating.

The frequency-weighting networks add components on a power basis; i.e., two equal components add 3 dB to the total, and a single component 10 dB greater than its neighbors essentially determines the level. Both of these methods (tangent-to-curve and frequency-weighting network) are very susceptible to tonal or narrow-band high-energy components. And it can be inferred from Licklider and Guttman[23] that tonal components do not mask speech very effectively.

Unlike the tangent and network methods—which are determined by one (tangent) or more (network) energy peaks and give readings equal to (tangent), or greater than (network), the highest peak—the SIL method lowers the importance of a peak by averaging in lower levels.

The fact that speech-intelligibility prediction is better when the 300–600-cps octave is included in the SIL calculation is not a new finding. Beranek[11] states on page 419, "... if the level in the 300 to 600 cps band is not more than 10 dB above that in the 600 to 1200 cps band, use the 600 to 1200 cps band as the first band and then define the *speech-interference level as the arithmetic average of the sound pressure levels in the three bands* 600 *to* 1200, 1200 *to* 2400, *and* 2400 *to* 4800 cps. However, if the levels in the 300 to 600 cps band are more than 10 dB above those in the 600 to 1200 cps band, the average of the levels in the *four* bands between 300 and 4800 cps should be used instead." Ship noises tend to predominate in low-frequency sound so it is not surprising that Strasberg[12] standardized on a 300–4800-cps SIL band when treating ship noises. The present data tend to support the view that the top octave 2400–4800 cps can be eliminated from the SIL calculations without undue loss in predictive ability. In fact, for these ship sounds the 2400–4800-cps octave just adds additional dispersion in the measurements.

SUMMARY AND CONCLUSIONS

Recordings of ship noises, office noises, and laboratory (shaped thermal) noises were assembled to make up a sample of 16 diverse noises. These noises were adjusted in over-all levels so that listeners hearing monosyllabic words at a constant level via a loudspeaker achieved 50% word-intelligibility scores. Insofar as it is possible to generalize from measurements based on 16 noise samples, physical measurements and calculations on equally speech-interfering noises show that the speech-interfering properties of noises is best estimated by averaging the sound-pressure levels in mid-frequency octaves, then by use of frequency-weighting networks in sound-level meters, and finally by fitting the spectral peaks to noise contours. More specifically: the best of the SIL methods, the 300–2400-cps SIL calculation, gave slightly better predictions than any other combination of octaves. In line with standardizing on "preferred frequency" octaves it is encouraging that the 3-band (500–2000 cps) SIL and the 4-band (250–2000 cps) SIL gave results similar to those obtained with the 300–2400-cps SIL.

Of the next best method (weighting networks), the A and DIN 3 were equivalent and as good as the 600–4800-cps SIL, all of which were better than the B and C networks. As a first approximation, and in the absence of octave band filters, A or DIN 3 sound-level meter measures may be used to predict the speech interference of steady-state noises.

Among the generally poorest methods (noise contour criteria), the ISO(R) was as good as the A or DIN 3-weighting network, or the 600–4800-cps SIL, and although the NCA curves gave better prediction than the ISO contours and the NC curves the tangent-to-curve methods as presently used are not very good, no better than the B-weighting network.

ACKNOWLEDGMENTS

We gratefully acknowledge the many people who assisted us in this study: D. D. Washburn, who assisted in the shipboard collection of noises; T. H. Wells, who assisted in the analysis and editing of the noises; W. E. Green, who was responsible for running the experiment to adjust the levels of noise to make them equally speech-interfering; J. L. Leonard, who performed many of the physical measurements on the equally speech-interfering noises; and R. S. Gales and R. W. Young, who critically read and offered constructive criticism on this manuscript. To all those and many others inside and outside of the U. S. Navy Electronics Laboratory, we offer our sincere thanks.

---

[23] J. C. R. Licklider and N. Guttman, "Masking of Speech by Line-Spectrum Interference," J. Acoust. Soc. Am. 29, 287–296 (1957). Data on Fig. 11 (p. 293) show that for 50% word articulation scores four sinusoids would need be 15 dB more intense than random noise. It should follow then that one or two sinusoids 10 dB above a random noise background would not further decrease word articulation scores.

Reprinted from SOUND, Vol. 2, No. 4, 25-29, July-August 1963
Printed in U.S.A.

Observer Variability in Reading Noise Levels with Meters

Roy G. Klumpp and John L. Leonard

LISTENING DIVISION, U. S. NAVY ELECTRONICS LABORATORY

San Diego 52, California

A VARIETY of instruments including oscilloscopes, power level recorders, thermocouples, and moving coil meters are available for measuring noise levels, but, because of factors such as low cost and ease of usage, in actual practice most noise level measurements are made with some kind of a moving coil, averaging meter calibrated to read in terms of root-mean-square (rms) values. Moving coil instruments in widespread use in noise measurements such as the sound-level meter, the VU (volume unit) meter, and the vacuum-tube voltmeter display signal level in terms of the position of a pointer against a calibrated scale. The level is read by an observer or observers and is usually recorded as a single number representing the average pointer position during the sampling period.

In routine noise-level measurements, it is not uncommon for two observers, both reading the same meter, to come up with two different numbers to represent the average level. If the noise being measured is relatively steady in level or fluctuates regularly within a limited range of levels, there is usually little disagreement. However, if the noise fluctuates markedly and irregularly in level, the job of estimating the average level is not easy, judgments are not made with confidence, and disagreement among observers is likely.

This problem is sometimes handled by averaging the estimations made by several observers, but, often, to save time, one person is appointed as the official meter reader, and his estimations are accepted as the correct value. The inaccuracy which may be introduced by this latter procedure is ordinarily ignored in routine noise measurements and is usually assumed to be less than the inaccuracy inherent in the measuring instrumentation or sampling procedures.

In some of our work connected with the measurement of the speech interference of noise, a number of noise levels were to be measured with sound-level meters and vacuum-tube voltmeters. It was not practical to use a group of observers, and it, therefore, became desirable to assess the bias which might be introduced by the use of a single meter reader. Accordingly an experiment was conducted to determine the dispersion among observers estimating the average level of samples of steady state and fluctuating noises. Each of nine observers estimated the average level of each of 16 different noise samples displayed on three different moving-coil meters.

Roy G. Klumpp received his B.S. in psychology in 1950 and his M.S. in the same field in 1951 from the University of Wisconsin. He joined the staff of the U. S. Navy Electronics Laboratory in 1951 as a Research Psychologist specializing in acoustics in air, particularly problems of hearing. Except for a two-year interruption to serve in the Army he has remained at NEL ever since. He has been author and co-author of numerous reports and articles in the field of acoustics. He is an associate member of the Acoustical Society of America. Since September 1962 he has been on extended educational leave working for his Ph.D. at the University of Iowa.

Mr. John L. Leonard received his B.A. in physics from the University of California at Berkeley in 1937. During World War II he was an Associate Physicist with the University of California. As a physicist for the Navy Electronics Laboratory, he has done work with underwater sound array models. For six years he headed an instrumentation staff technically responsible for all instrumentation and physical measurements for research in the Human Factors Division at NEL. He has been author and co-author of many reports and several articles and holds several patents.

Procedure

From among a number of meters on hand in the laboratory, three meters were chosen, primarily on the basis of a difference in damping characteristics. The three meters selected were: (1) a Hewlett-Packard Company, model 403A transistor voltmeter; (2) a General Radio Company, type 1551-A sound-level meter; and (3) a Bruel and Kjaer, model 2409 vacuum-tube voltmeter. Figures 1, 2 and 3 are photographs of the three meters, and

TABLE I. Characteristics of the meters as used in the experiment.

Meter	Dial settings	Range of scale markings in dB	Scale divisions per dB	Frequency response	Approximate time in sec to reach 100% scale
Hewlett-Packard 403A Voltmeter	"1 cps-1 Mc"	−12 to +2	1	Uniform	8.0
General Radio 1551A Sound-Level Meter	"A Weighting" "Slow Meter"	−6 to +10	1	Reduced response below 1000 cps	3.5
Bruel & Kjaer 2409 Voltmeter	"Peak," "VU Damping"	0 to +20	1, 2	Uniform	0.3

Table I lists several characteristics of the meters as they were employed in the experiment.

It is apparent from Figs. 1–3 and Table I that the meters differed in a number of respects. To the observer reading the meters in this experiment, however, the primary difference among the meters was the degree of damping. The Hewlett-Packard (HP) meter set for "1 cps-1 Mc" (cycles per second and megacycles per second) response almost completely "ironed out" momentary fluctuations in level; the Bruel and Kjaer (B & K) meter set to "Peak-VU" readily responded to most momentary fluctuations; the General Radio (GR) meter set to "A weighting-Slow Meter" was intermediate among the three meters in the rapidity of its response. Differences in damping among the three meters are shown in Fig. 4, which plots the per cent of pointer angular deflection as a function of time when each was activated by the abrupt onset of a sustained 1000-cps signal. Figure 4 shows that, as the meters were adjusted, the time to reach 70% of full-scale angular deflection was less than 0.2 second (sec) for the B & K meter, about 0.5 second for the GR meter, and over 2.5 seconds for the HP meter. The curves shown were obtained using an observer with a stop clock to time pointer angular deflections, and are, therefore, approximations.

Table II presents information on the 16 noise samples used in the experiment. Some of the noise samples were recorded aboard operating Navy ships, some were recorded in shops and offices on land, and a few were generated in the laboratory. The noise samples in Table II are ordered from 1 to 16 on the basis of frequency characteristics. Noises No. 1 through No. 5 contain predominantly low frequencies; noises No. 6 through No. 12 have most of their energy in the mid-frequency region, and noises No. 13 through No. 16 have strong high-frequency components.

The duration of each sample, listed in column 4 of Table II, was tailored to be sufficient for an observer to make a single judgment on a meter. Noise samples which varied only slightly or regularly in level were presented for a minimum length of time (9 sec). Other samples which varied irregularly or widely in level were presented for periods of time up to 44 sec. A 5-sec interval between samples permitted observers to record estimations made on the meters.

The observed variation in level of each noise sample differed among the three meters because of differences in meter damping, frequency response and

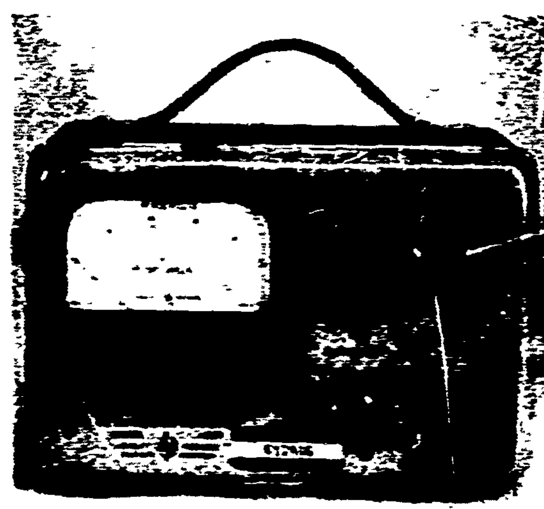


Fig. 1. Hewlett-Packard model 403A Voltmeter.

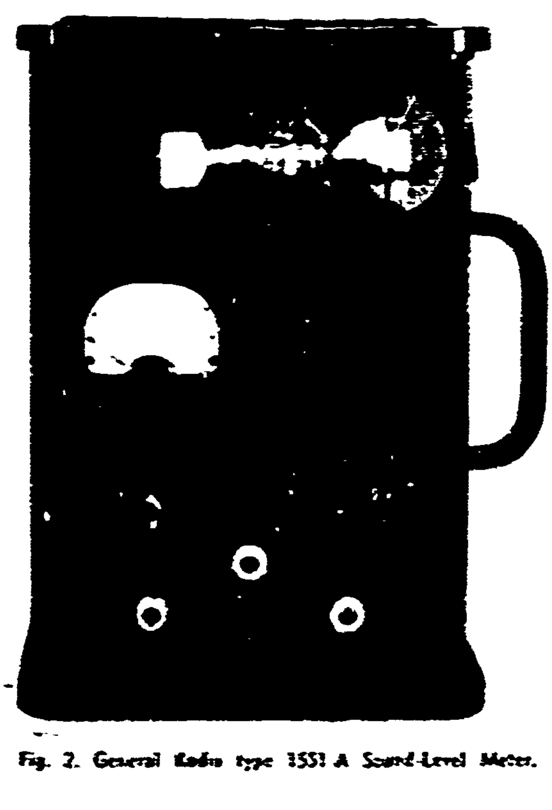


Fig. 2. General Radio type 1551 A Sound-Level Meter.

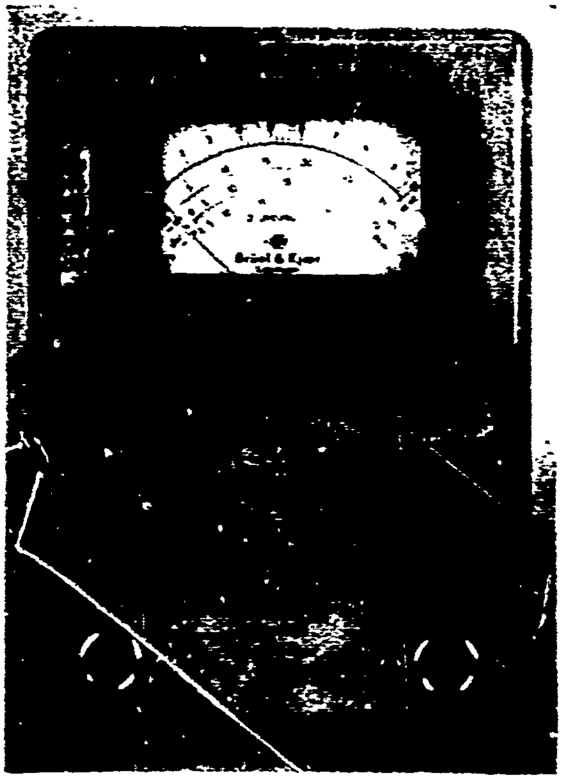


Fig. 3. Bruel and Kjaer model 2409 Voltmeter.

rectified circuitry. Figure 5 shows noise traces made with a power-level recorder adjusted to simulate the response of the B & K meter, i.e., VU damping, peak reading. This figure shows that a few of the noises varied relatively little in level, some varied moderately, and a few varied markedly in level. Each trace represents a time interval of slightly more than 15 sec, so that one complete cycle of the longest noise sample (11) could be displayed.

The meters were connected to the output of a tape reproducer with the signal fed into the GR sound-level meter via the microphone connector. The recorded level of each of the noise samples was as far as possible adjusted to allow the 16 samples

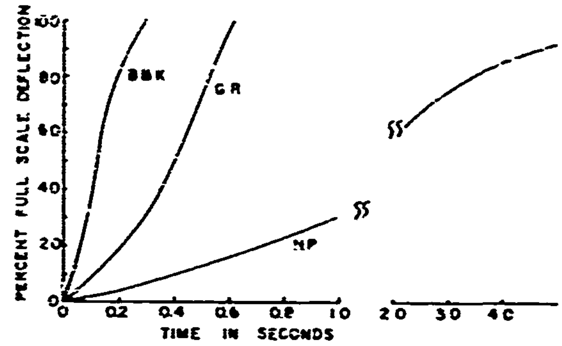


Fig. 4. Approximate damping characteristics of the three meters used.

to be displayed without changing meter settings between samples. Exceptions were noises 1 and 2 for which the gain of the GR meter was increased by 20 and 10 dB (decibels), respectively, to compensate for the de-emphasis of low frequencies in this meter set to "A" weighting. No attempt was made to make the three meters agree on any single noise or set of noises. The meters were, in fact, adjusted to read differently, so as to minimize the possibility of estimations made on one meter influencing estimations made on another meter.

To enable subjects to hear the noise at the same time they were reading the level on a meter, the noise samples were reproduced over an extended range loudspeaker system at a comfortably loud level. All subjects were employees of the Navy Electronics Laboratory, and all had considerable experience in reading many types of meters.

Each subject served in three test sessions—one on each of three days. During each test session each subject made an estimation of the average level of each of the noise samples on each of the three meters (16 noises times 3 meters equals 48 judgments/session). The sequence of meter presentation versus subjects was counterbalanced over the three sessions.

Preparatory to the first test session, the noise samples were played through twice to acquaint the subjects with the noises, meters, and test proce-

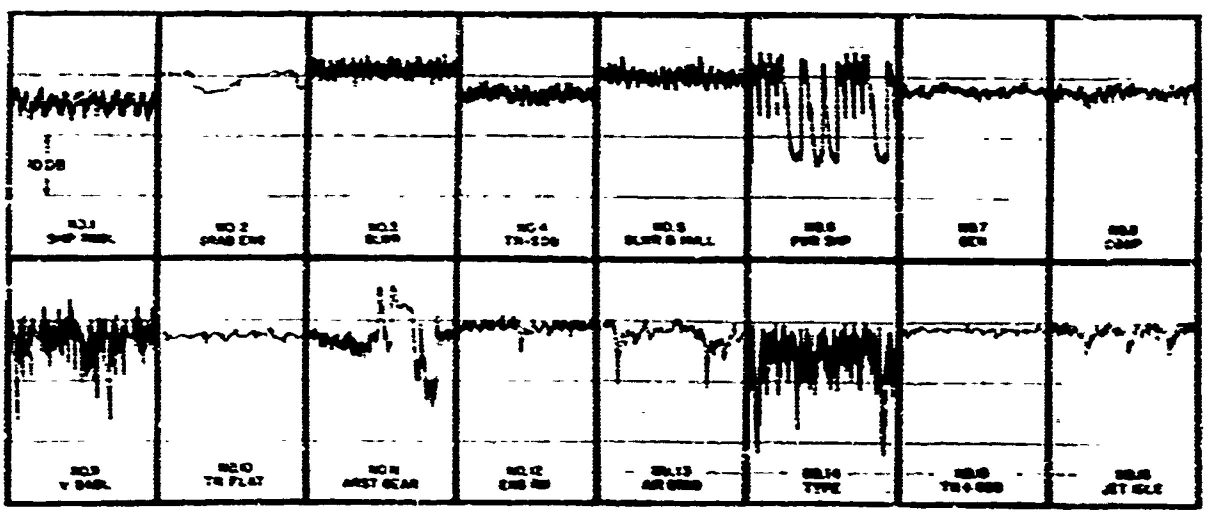


Fig. 5. Power level tracings of the 16 noise samples.

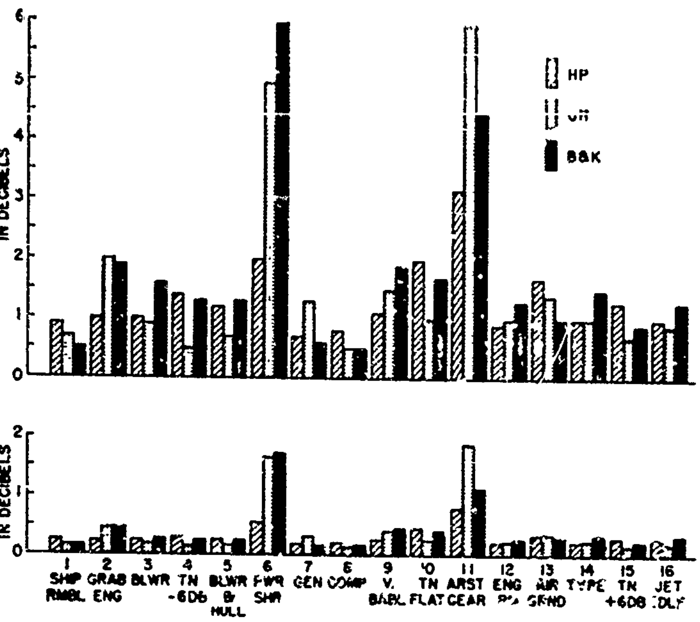


Fig. 6. Standard deviations and ranges of estimations made by 9 observers on each of three days.

dures. In each test session the subject was instructed to write down for each noise sample a single number, representing his best estimate of the average level of the sample. Subjects were told to observe the pointer position during the entire time the sample was being presented and to obtain estimations by using an "arithmetic-averaging" procedure. Thus, according to instructions, if the pointer deflection was "2" for ½ the time and "8" for ½ the time, the desired single-number average reading would be "5.0." If the deflection was "2" for 10% of the time and "8" for 90% of the time, the desired reading would be "7.4."

It was apparent that the subjects would not be able to perform computations like those illustrated in the limited time available during and between noise samples. The instructions about averaging were given to insure that all subjects started with the same orientation and in the hope that, although specific instructions could not be followed, the general philosophy would be adhered to.

Subjects were instructed to read only the decibel scales on the meters and to make estimations to the nearest 0.1 dB whenever possible. Only numbers corresponding to those on the meter scales were recorded, since, with the meters deliberately thrown out of calibration, the absolute values had no meaning. Before each of the test sessions, the stability of the playback and meter system was checked. Changes in the level of recorded sine waves and constant level noise signals were less than ± 0.2 dB as measured with a single observer reading a precision voltmeter.

Results

Figure 6 shows, in histogram form, the standard deviation and range of estimations for each of 16 noises made on each of three meters. Each bar is based on 27 judgments (one judgment on each of three days from each of nine subjects). The numbers and abbreviations across the bottom of the figure identify the noise samples (see Table II). The bars for each noise are arranged with the most highly damped meter (HP) to the left, the intermediately damped meter (GR) in the center, and the least damped meter (B & K) to the right.

Figure 6 shows that for 14 of the 16 samples (5 and 11 excluded) the standard deviation of estimation is less than 0.5 dB and the range is 2 dB or less.

An analysis of variance was made on the data obtained with all 16 noises and also with the data from 14 noises (6 and 11 excluded). Results of the two analyses at the 1% level of confidence were the same. Although all main variables (Noises, Meters, Days, and Subjects) were significant at the 1% level, an examination of the standard deviations shown in Fig. 6 indicates that the total variability from trials and subjects is in actuality very small.

Discussion

For 14 of the 16 noises the range of estimations is 2 dB or less and the standard deviation is less than 0.5 dB. Considering that several of the noise samples contain wide variations in level, the agreement among 27 estimates (nine subjects on three different days) is surprisingly good.

Although the three meters differed in a host of characteristics which might be expected to affect the estimations (damping, frequency response, arrangement of scales, size, and number of scale divisions, etc.), there appears to be no clear-cut advantage in using one meter in preference to another except for noises 6 and 11. For these two noises the heavily damped meter (HP) produced the smallest dispersion of estimations.

The obtained standard deviations and ranges on noises 6 and 11 were two to three times larger than for the other 14 noises. Clearly the ability of the subjects to agree on time-level estimations is limited. It should be noted that strong objections were voiced by most of the subjects about making the required single-number estimation on noises 6 and 11. They felt that these two noise samples were made up of several distinctly different components and should, therefore, have been estimated in terms of two or more separate levels. Sample 6, the power shear, consisted of a number of short sharp "bangs" in a sustained, low-frequency background of motor noise while sample 11, the arresting gear noise, consisted of low-frequency motor and ship noise, a shriek changing in amplitude and frequency, and a single short duration "bang." Subjects felt they were able to estimate the level of each component in these noises with a reasonable degree of accuracy but could not assess accurately the relative duration of momentary components such as "bangs" and "shrieks."

In the present study subjects were instructed to estimate the average noise level by averaging the pointer deflections on a time-weighted basis. This procedure was used, because the instructions are relatively easy to follow and the procedure is, in fact, one that is commonly used in noise-level reading. It should be recognized that such an averaging procedure does not result in an accurate estimation of the true sound-pressure level, especially when the levels being combined differ substantially.

Measurements of the level of a noise made with a moving-coil meter are subject to error from at least three sources: (1) the meter, itself; (2) the measurement technique; and (3) variability or bias from the observer. It is of interest to speculate about the probable magnitude of these errors. It is difficult to estimate the error attributable to meters because of differences in the precision and circuitry of meters in current use and because the error in a moving coil meter is dependent upon the characteristics of the noise being measured. If the desired quantity is the true rms level, it is unlikely that commercially available meters in ordinary use will be more accurate than ± 0.5 dB and may well be from 1 to 5 dB or more in error on certain noises.[1] Errors attributable to measurement techniques are also not easily determined. In the measurement of noise in electrical systems, the sampling procedure is usually not critical. However, in measurements made on acoustic signals, an optimum sampling procedure cannot always be clearly specified or followed, and, in such cases, the measurement technique error can easily be several times the minimum error attributable to the measuring instrument.

The data in this experiment suggest that the error contributed by the meter observer, provided the averaging procedure is specified, is small enough to ignore in routine noise measurements.

1. Moving coil meters with special rectifying circuitry are available for measuring the rms level of noise. Provided the noise being measured does not have an extreme peak to rms ratio or is not extremely short in duration, such meters are accurate to within perhaps ±0.5 dB. The conventional voltmeter and VU meter, calibrated in terms of the rms of a sine wave but actually responding to the average value of signals, may deviate markedly from true rms when used to measure noise signals. See B. M. Oliver, Hewlett-Packard J. **12**, No. 7, 1–4 (1961); E. E. Gross, General Radio Company Experimenter **32**, No. 17, 3–9 (1958); C. G. Wahrman, Bruel and Kjaer Tech. Rev. No. 3, 9–21 (1958); James J. Davidson, "Average vs. RMS meters for Measuring Noise," IRE Trans. on Audio, No. 4 (1961) and F. E. Terman and J. M. Pettit, *Electronic Measurements* (McGraw-Hill Book Company, Inc., New York, 1952), 2nd ed., Chap. 1.

TABLE II. Description of the 16 noise samples.

Name	Description	Maximum change in level in dB (from Fig. 5)	Duration of sample in sec
1. Ship Rumble	Primarily very low-frequency ship rumble	6	15
2. Grab engine	Electrical motor of large size with strong low-frequency hum-like components	4	35
3. Blower	Low-frequency blower noise recorded aboard ship	6	9
4. Thermal noise, −6 dB per octave slope	Low frequencies predominant	4	10
5. Blower and hull	Low-frequency rumble and machine noise	6	10
6. Power shear	Hydraulic power, shear cutting metal, includes clang of sheet metal dropping to floor	19	20
7. Generator	Shipboard generator with whine and low-frequency components	4	26
8. Compressor	Refrigerator compressor with rhythmic sound	5	15
9. Voice babble	5 pairs of communicators exchanging monosyllabic words	21	31
10. Thermal noise, flat	Wide-band hiss	3	9
11. Arresting gear	Noise recorded in arresting gear room of aircraft carrier has rumble, shriek, bang, and roar components	21	44
12. Engine room	Complex machinery noise wide band	6	12
13. Air grinder	Air-driven grinder on metal sheet; high pitched "dental drill" sound	11	20
14. Typewriter	Typewriter operated at 40–60 words per minute—strong in high frequency components	22	28
15. Thermal noise +6 dB per octabe slope	High-frequency hiss	3	9
16. Jet idle	Jet engine on flight deck of carrier—has strong whine components	7	20

A CONTRIBUTION OF THE NAVY ELECTRONICS LABORATORY, SAN DIEGO 52, CALIF[illegible]

Reprinted from THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, Vol. 35, No. 9, 1339-1344, September 1963

Articulation Index and Average Curve-Fitting Methods of Predicting Speech Interference

J. C. WEBSTER AND R. G. KLUMPP

U. S. Navy Electronics Laboratory, San Diego, California
(Received 5 March 1963)

Sixteen equally speech-interfering noises were rated by using articulation index (AI) calculations and by using noise criteria (NC)-type contours as averaging, not peak-finding, devices. Articulation-index scores based on 5 or 6 weighted octave-band levels were equal to AI scores obtained by the 20-band method and predicted very well that the noises were equally speech-interfering. The use of NC-type curves (from 500 to 2000 cps) to find an average, not a peak, noise level also gave good prediction. Equally speech-interfering noises were not, however, equally "loud," nor equally "noisy." There was evidence to show that the frequency that divided noise-masked speech into two equally intelligible frequency regions was 850 or 1000 cps, and was not 1700 or 1900 cps, which is the dividing frequency when the speech in quiet is progressively restricted in bandwidth by high- or low-passed filtering.

INTRODUCTION

IN two earlier papers,[1,2] methods of measuring the speech-interference properties of 16 diverse noises were discussed. The general procedure was that 16 noises were adjusted in level so that listeners hearing monosyllabic (rhyme) words at a constant level via a loudspeaker obtained 50% word-intelligibility scores. Then various physical or psychophysical measurements were made on the 16 noises reproduced at the equally speech-interfering levels. Many measurements and/or calculations were not reported in the two previous papers because (1) only limited presentation time and space were available,[1] and (2) only simple schemes were under study.[2]

This paper will deal with the articulation index (AI), the use of tangent-to-curve contours as an averaging method, and will discuss the common spectral characteristics of the noises.

I. ARTICULATION INDEX

The AI involves the difference in levels between the speech spectrum and the various noise spectra in different bandwidths. This S–N difference in dB can range from zero to 30, since no negative values are allowed, and all values greater than 30 are called 30. S–N, from 200 to 6800 cps, can be calculated in 20 bands, varying in bandwidth; or in third-octave bands; or in octave bands. Regardless of how many bands are used, an average S–N in dB (between 0 and 30) is found and divided by 30 to get a number from 0 to 1 which is called the articulation index. For further background information on AI, see Licklider and Miller[3] for history and rationale, Hawley and Kryter[4] for details of its use, Licklider[5] for a critical review of the assumptions involved, and Kryter[6,7] for recent modifications, refinements, and revalidation.

In this paper it will be desirable to compare the AI results to measures from Refs. 1 and 2 always meas-

[1] R. G. Klumpp and J. C. Webster, "Predicting Speech Interference from Physical and Psychophysical Measures of Ambient Noise," J. Acoust. Soc. Am. 35, 1116(A) (1963).

[2] R. G. Klumpp and J. C. Webster, "Physical Measurements of Equally Speech-Interfering Navy Noises," J. Acoust Soc. Am. 35, 1328 (1963).

[3] J. C. R. Licklider and G. A. Miller, in *Handbook of Experimental Psychology*, edited by S. S. Stevens (John Wiley & Sons, Inc., New York, 1951), Chap 26, pp. 1055–1058.

[4] M. E. Hawley and K. D. Kryter, "Effects of Noise on Speech," in *Handbook of Noise Control*, edited by C. M. Harris (McGraw-Hill Book Company, Inc., New York, 1957), Chap. 9, pp. 9–5 to 9–11.

[5] J. C. R. Licklider, "Three Auditory Theories," in *Psychology: A Study of a Science. Study I, Conceptual and Systematic. Vol. I: Sensory, Perceptual and Physiological Formulation*, edited by Koch (McGraw-Hill Book Company, Inc., New York, 1959).

[6] K. D. Kryter, "Methods for the Calculation and Use of the Articulation Index," J. Acoust. Soc. Am. 34, 1689–1697 (1962).

[7] K. D. Kryter, "Validation of the Articulation Index," J. Acoust. Soc. Am. 34, 1698–1702 (1962).

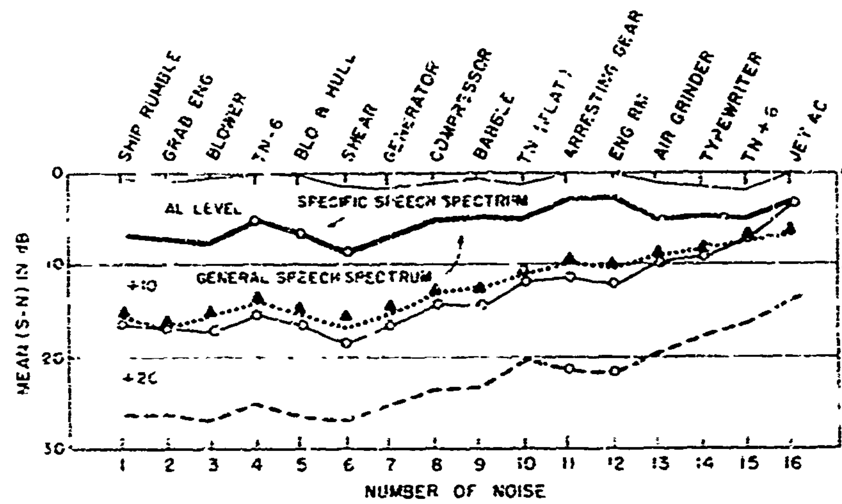


FIG. 1. Articulation index (AI) calculations for 16 equally speech-interfering noises. The ordinate is the AI (right) or the average speech-to-noise difference over 20 equally important speech bands or over 5 to 6 octaves (left). The ordinate increases in a positive direction downward so the AI data will be compatible with all other data (SIL, weighting network, etc.) measured in this paper and in Ref. 2. That is, the higher the data points, the noisier the noise is rated.

Four different speech levels were used to make the AI calculations, the actual level used and levels at −10, +10, and +20. Open circles mean that S–N levels greater than 30 or less than zero never occurred. All lines are AI figured on the 20-band method. The shading around the "actual level" line indicates the difference between using the actual or specific speech spectrum or a generalized speech spectra both on the 6-octave weighted method of calculation (from Ref. 2). The data points around the +10 line indicate the 5-octave (Δ) or the 6-octave (·····) method of calculation at a speech level of +10. The shaded area at the right end of the +10 line shows the difference between limiting the AI speech-to-noise ratios to between 0 and 30 (lower boundary) or letting the speech-to-noise ratios take on any + or − value (open circles and upper edge of boundary).

ured in dB, so the AI will be left in terms of average S–N. Although it is not always possible to compare AI's directly to SIL's, it is possible in this study because the speech level was constant for all 16 noises, the distance between loudspeaker and listener was fixed, and all listening was done in the same room. Since the speech level and speech spectrum were always the same, the AI for each noise depends on the noise spectrum and level just as the SIL does.

Figure 1 shows the results of AI calculations using the 16 noise spectra from Figs. 1 through 16 in Ref. 2. The speech spectrum used for all but one calculation is the "general speech spectrum" from Ref. 6. The solid and dashed lines show AI's calculated on the 20-band method, using the octave-band spectral data from Ref. 2 as the basis for the noise spectra, and using for the speech the "general speech spectrum." The heavy solid line between 0 and 10 dB (or 0 and 0.33 AI) is the AI calculation based on the "actual level" of speech used in this experiment. The lighter lines are for speech levels ±10 dB from "actual," and the dashed line is for an assumed speech level +20 dB above the level actually used.

The shaded area around the "actual level" line shows what the AI is when the 6-octave-band method is used instead of the 20-band method. The upper edge is when the "specific speech spectrum" and the "actual level" as used in this experiment are used. The lower boundary is when the "general speech spectrum" is used with the 6-octave method.

The triangles and the dotted line around the "+10" speech-level line are for the 5-octave and 6-octave calculation schemes, respectively.

A circle on any of the lines means that in none of the 20 bands did the S–N truncate at 0 or 30 dB. If each line had nothing but these circles, the lines would be exactly 10 dB apart. Because of these truncations, however, on very few noises are the lines exactly 10 dB apart. There is no line that is free of truncations completely. The "+10" line truncates only on noises 13 through 16, and for these noises the true (truncated) AI values are shown as the bottom edge of the shaded portion. The line and circles for noises 13 through 16 on the "+10" line are the values that the AI would have assumed if the 0- to 30-dB limitations had not been adhered to.

The noises in this experiment were adjusted in level to limit listeners' scores on rhyme words to 50%. The average AI for a 50% rhyme score is 5.4 dB or 0.18 (averaged over the 16 noises on the "actual level" line of Fig. 1). Montague[8] shows that a 50% rhyme score is equivalent to a 40% PB word score, and Kryter[6] finds that a 40% PB score is predicted by an AI of 0.17. At this low differential level of speech to noise (5.4 dB), many bands (of the 20) on many noises (of the 16) show negative S–N differences. These negative values are called zero, and, because of the preponderance of these truncated bands at the zero level, the "actual level" line is quite horizontal in appearance. Actually, on only

[8] W. E. Montague, "A Comparison of Five Intelligibility Tests for Voice Communication Systems." Navy Electronics Laboratory Report 977. PB 157-229. AD 254-545 (1960).

3 of the 16 noises (those indicated by the open circles) do all values for all 20 bands fall above zero (and less than 30). All other noises truncate on at least one band.

It is only when a speech level 10 dB greater than that actually used is the basis of the calculation that truncation within the 0- to 30-dB range is largely avoided. When a speech level 20 dB greater than the level actually used in these experiments is used as the basis of calculation, truncations occur when S–N values of greater than 30 must be called 30. This occurs on all noises except noises 11 and 12.

In summary, in regard to speech level: There is no level of speech (in the 10 dB steps chosen in this study) where all 20 bands in all 16 noises are within the 0- to 30-dB acceptable range. Using the "+10" level of speech and allowing the values to go negative (or exceed 30), a rating of the noises that is independent of the 0- to 30-dB range results. This is no longer a true AI calculation, but it does give a speech-interference measure on a noise spectrum by utilizing the speech spectrum. The true AI takes into account both the speech spectrum and the 0- to 30-dB S–N range. If the speech level is progressively decreased, the AI rates the noises progressively more equal in speech interference until, of course, all noises mask out speech completely, at which point all noises are indeed equally speech-interfering.

None of the calculations shown in Fig. 1 include any of the spread of masking corrections detailed by Kryter.[6] This is because at the levels of noise used in these calculations there is essentially no spread of masking. For example, on noises 1 and 2 the maximum level in the octave around 200 cps is 90 dB. When this is converted to spectrum level, it is reduced to 67 dB and the upward spread of masking is for only 150 cps, and then it falls off at 25 dB per octave (Table IV, Ref. 6). This correction would at most affect only the lowest of the 20 bands and then by less than 5 dB. The total correction in the 20-band average would then be a quarter of a dB which is not distinguishable on plots such as Fig. 1.

However, the AI, even without the correction for upward masking, does a good job of specifying that these 16 noises, adjusted in level to be equally speech-interfering, are indeed equally speech-interfering. Although no table of scores and standard deviations will be included in this paper, as was done in Ref. 2, the standard deviation of the AI calculated by the weighted octave-band-level method and using the "general speech spectrum" was 2.4 based on the 16 noises (or 2.1 using the 33 weighting of Ref. 2). This makes the AI almost identical to the best rating method, the 300–2400-cps speech-interference level (SIL) of Ref. 2. And when the 20-band AI of the "specific speech spectrum" was used, the AI had a slightly lower standard deviation; namely, 1.7.

The data in Fig. 1, therefore, tend to support Kryter's[7] latest evidence in that they "... amply demonstrate the general validity of the AI calculated by the 20-band method...." The data also tend to support the conclusion of Kryter, Flanagan, and Williams[9] that "the octave band method for the calculation of AI can be used in place of the more detailed 20-band method without any appreciable loss in the accuracy with which speech intelligibility test scores are predicted." It would also appear that, for comparative purposes, a general speech spectrum is nearly as useful as the specific speech spectrum in calculating the AI.

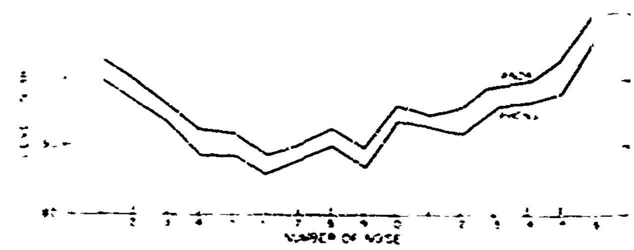


FIG. 2. Loudness (in phons) and noisiness (in PNDB) of 16 equally speech-interfering noises.

II. "LOUDNESS" AND "NOISINESS"

Are noises that are equally speech-interfering equally loud or equally noisy? The literature abounds in methods of calculating loudness,[10,11] and a concept of perceived noisiness,[12] measured in noys and possibly related to annoyance, has recently been formulated. The loudness and noisiness of these 16 equally speech-interfering noises have been determined by use of the Mark 6 loudness contours of Stevens,[10] and Kryter's revised noys contours,[13] and are plotted in Fig. 2.

The curves in Fig. 2 show that equally speech-interfering noises are neither equally loud nor equally noisy. In fact, the relative rankings among the 16 noises are rated by loudness or noys methods almost as they were in Ref. 2 by the NCA curves.[14] The 16 noises are, however, noted to be about 12 dB higher in phons and 15 dB higher in PNDB than by the NCA contours.

The NCA rating is determined by the highest peak in a noise spectrum that touches an NCA contour. Both the phon and PNDB calculations give maximum weight to the highest noise peak but also add in a fraction of all lower levels all along the frequency spectrum. It might be expected therefore that the loudness or noisiness of these 16 equally speech-interfering noises would be more nearly equal than the NCA ratings because

[9] K. D. Kryter, G. Flanagan, and C. Williams, "A Test of the 20 Band and Octave Band Methods of Computing the Articulation Index," Bolt Beranek and Newman Inc., Contr. USAF19(604)-4061, Rept. ESD-TDR-62-4 (1961).

[10] S. S. Stevens, "Procedure for Calculating Loudness: Mark VI," J. Acoust. Soc. Am. 33, 1577–1585 (1961).

[11] E. Zwicker, "Ein Verfahren zur Berechnung der Lautstärke," Acustica 10, 304–308 (1960).

[12] K. D. Kryter, "The Meaning and Measurement of Perceived Noise Level," Noise Control 6, No. 5, 12–27 (1960).

[13] K. D. Kryter (personal communication); also shown as an appendix (Fig. A.1) in J. T. Broch, "Loudness Evaluation," Bruel & Kjaer Tech. Rev. No. 2 (1962).

[14] L. L. Beranek, "Revised Criteria for Noise in Buildings," Noise Control 3, No. 1, 19–27 (1957).

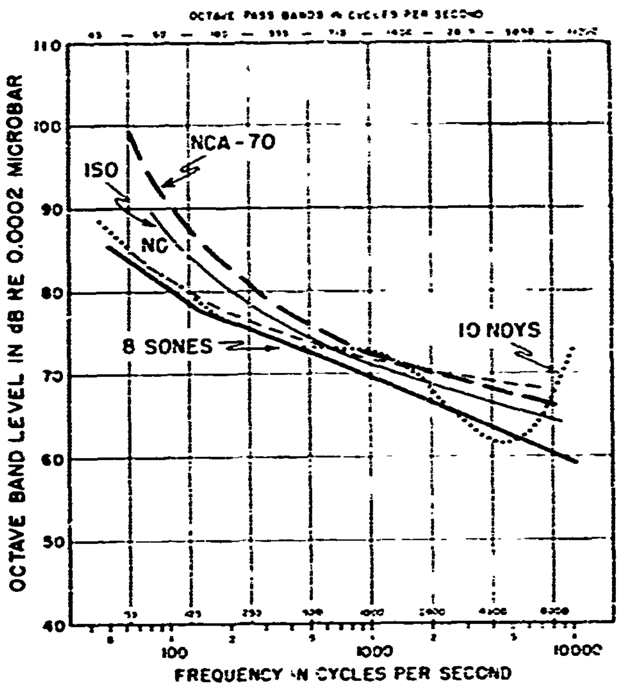


FIG. 3. Plot of the NCA-70, ISO-70, NC-70, 10 noys and 8 sone contours.

some account is taken of noise levels other than the highest peak. And, in fact, there is a slight reduction in the dispersion of scores: namely, from a standard deviation of 5.2 for the NCA rating (Table 2, Ref. 2) to 4.8 for both the phon and PNDB curves. However, this small amount of leveling does not appreciably change the shape. It still must be stated that rating noises by undue regard to the highest peak that touches a family of noise- (or loudness-) rating contours is not a good method of describing the speech-interfering properties of said noises.

III. AVERAGE CURVE-FITTING TECHNIQUES

In Ref. 2 it was observed that fitting the peaks of plotted noise spectra to families of NC,[14] NCA,[14] and ISO[15] curves was not a good way to rate these noises as being equally speech-interfering. And in Sec. II of this paper, it was observed that loudness[10] or noys[12] calculations were not much better. However, it has been shown[1,2] that by using only the portion of the (ISO) curves that were in the speech region, say, those octaves centered at 500, 1000, and 2000 cps, a great improvement resulted.

It should also be noted that if only those parts of the contours centered on the octaves at 500, 1000, and 2000 cps are used, the families of contours are not radically different. Note, for example, in Fig. 3, that the contours of the NCA-70, ISO-70, NC-70, the 10 sone, and the 10 noys are all within 5 dB of each other through this range. And, in fact, the average values are for the NCA, 72; ISO, 71; NC, 72; 10-sone, 73; and 10-noys, 70 dB. It follows then that any of the above families of contours would give considerably better (lower dispersion) ratings on the 16 equally speech interfering noises if restricted to the 500-, 1000-, and 2000-cps range, and that any one curve would be nearly as good as any other. In Ref. 2 only the ISO was used in the restricted range, but any other one would have worked as well.

However, restricting the range on the contours while continuing to use the tangent-to-curve method does not describe the speech-interfering properties of the 16 noises as well as the 500-, 1000-, and 2000-cps SIL or the "actual level" AI.

It is possible that even better predictions could be made if, instead of fitting the peaks of noise spectra to families of curves, some sort of averaging or integration could take place. For example, separate readings can be made of where the ISO, NCA, or noys contours, at the points 500, 1000, and 2000 cps, become tangent to the noise spectra, and an average of these three readings taken; or another way of arriving at the same rating is by visual averaging. To accomplish this the curve, say, the ISO-70, is fitted over the 16 noise spectra such that the best visual average is obtained (half the spectrum lies above and half below the ISO-70 curve). When this average fit is obtained, the point where the 70-dB ordinate on the ISO curve intersects the ordinate on the noise spectra is the desired rating. This is roughly equivalent to making an SIL calculation; the difference is that the contours are not flat through these three octaves. It is therefore not surprising to note in Fig. 4 that the results of averaging the ISO-70 contour through the 16 noise spectra are virtually indistinguishable from the 3-band preferred-frequency SIL (500, 1000, and 2000 cps) replotted from Ref. 2. And the standard deviations are also equal at 2.8.

It would appear from these results that a better way to use NC, NCA, or ISO contours to rate the speech-interfering properties of noise would be to find the lowest noise-rating curve that *averages out* the variations in the noise spectrum at 500, 1000, and 2000 cps and

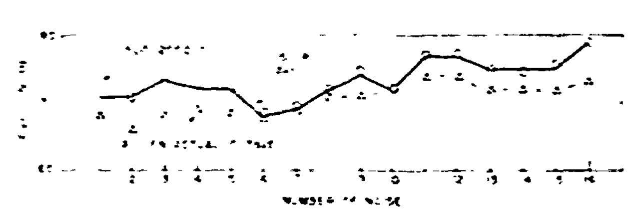


FIG. 4. Speech interference [ISO(R)AVE] based on averaging the midfrequency region (500, 1000, and 2000 cps) of the ISO contours through the 16 equally speech-interfering noise spectra. Plotted from Ref. 2 for comparison: the SIL (500-, 1000-, and 2000-cps 3-band average) as open circles, and as triangles the AI (general speech spectrum, actual level, and on octave-band basis) arbitrarily set equal on noise No. 10 (thermal, flat noise).

[14] These International Standards Organization curves appear in drafts of Tech. Comm. 43 (see footnote 17 in Ref. 2), but are also published with instructions on their use by J. H. Janssen, "Some Acoustical Properties of Ships with Respect to Noise Control. Part I," Report No. 41S of Netherlands' Research Centre T.N.O. for Shipbuilding and Navigation, Delft, The Netherlands (1962).

not, as presently used, determine the lowest noise-rating curve *just not exceeded by* the plotted noise spectrum. In other words, average out any noise peaks; don't let the highest noise peak determine the rating.

IV. COMMON SPECTRAL CHARACTERISTICS

Many of the results of this study point to a re-examination of the "most important speech frequency." For example, the SIL that predicted these data best was the 300-2400-cps average,[2] or the 3-band preferred-frequency octaves centered at 500, 1000, and 2000 cps, and not the more conventional 600–4800-cps average. Likewise, the AI weighted the low frequencies too little and the high frequencies too much. And when tangent-to-curve methods or average curve-fitting methods were restricted to the three octaves centered around 1000 cps, they predicted much better. These data tend to imply that the lower-frequency components of the masking noises are more important than other investigators have thought them to be.

For example: French and Steinberg,[16] using speech in the quiet and normal listeners, progressively high- and, later, low-pass filtered the speech until it became progressively less intelligible. They found that speech was equally deteriorated when all frequencies either above or below 1900 cps were filtered out; i.e., the frequency range above 1900 cps was as important as the frequency range below 1900 cps. Beranek,[17] using male voices only, found the crossover frequency to be 1700 under the same quiet-filtered-speech conditions.

However, Pollack[18] redid the filtered-speech intelligibility studies, but added a broadband noise background and varied the level of the speech. He found that the crossover (or equal importance) frequency increased from 800 cps for low levels of speech through 1010, 1300, 1430, to 1620 cps for increases of 10 dB in the speech level.

Dyer[19] did the reverse of Pollack: namely, he filtered the noise around a broadband speech signal and found, like Pollack, that as the speech-to-noise differential increased the crossover frequency increased from about 1000 cps to almost 2000 cps. A crossover frequency can also be found from Klumpp and Webster data[2] and is shown in Fig. 5. To arrive at Fig. 5, the spectrum of Fig. 10 (Ref. 2) was subtracted from the spectra of Figs. 1 to 9 and 11 to 15. This amounts to taking out the characteristics of the playback system by subtracting the flat thermal noise spectrum from all other spectra. This is roughly equivalent to measuring the electrical voltage across the transducer instead of the acoustical output. This procedure in no way changes any of the interrelationships among the noises; it only assigns a value of 0 dB in each octave to the flat noise spectrum and plots all others relative to this.

The noise spectra in Fig. 5 are plotted in groups: at the top are the envelopes of spectra of noises 1 through 5 and 12 (Figs. 1 through 5, and 12 of Ref. 2). That is, a curve is drawn through the highest level at each octave of any of the 6 noises. A similar curve is drawn through the lowest level assumed by any of these noises. The envelope at the top of Fig. 5 encompasses therefore the total range assumed by any of the low-frequency noises between the frequency limits of 250 and 2000 cps.

Similarly, the envelopes of noises 6, 7, and 8 are plotted and show a much smaller dispersion. The plotted envelopes of noises 9, 10, 11, 13, and 14 are slightly upsloped. Noise 15 is plotted by itself since it is the only noise which is predominantly a high-frequency noise. The jet noise is not plotted since it is unduly influenced by the single frequency components at 3000 cps and above.

At the bottom of Fig. 5 is the average spectrum of the three types of noise envelopes, together with the one predominantly high-frequency noise spectra. Observe in Fig. 5 that both the predominantly high- and low-frequency noise spectra and the slightly high- and low-frequency noise spectra cross each other at about 850 cps.

Kryter[7] also has some data of speech masked by noise which, when replotted, show a low crossover frequency. For example, if, instead of plotting the noise spectra of

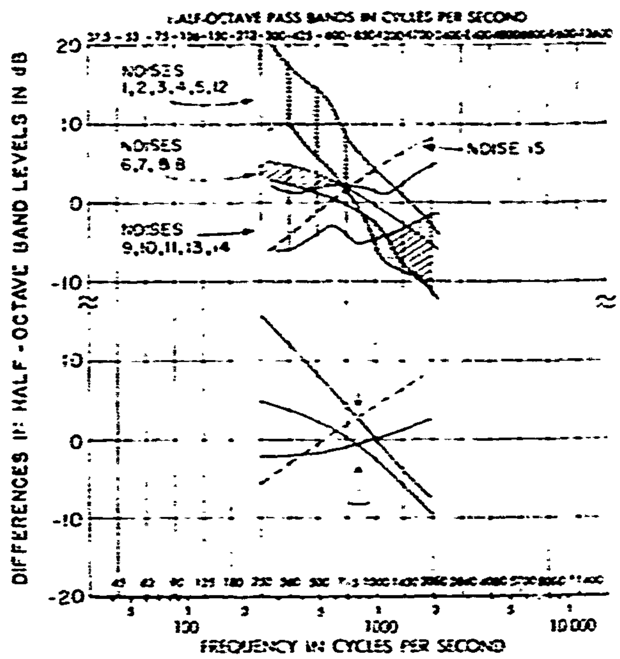


FIG. 5. Ranges (top) and averages (bottom) of 4 types of noise spectra all adjusted in level to be equally speech-interfering. (Add 3.1 dB to obtain octave-band level.)

---

[16] N. R. French and J. C. Steinberg. "Factors Governing the Intelligibility of Speech Sounds," J. Acoust. Soc. Am. 19, 90-119 (1947).

[17] L. L. Beranek, "Design of Speech Communication Systems," Proc. IRE 35, 880-890 (1947).

[18] I. Pollack, "Effects of High Pass and Low Pass Filtering on the Intelligibility of Speech in Noise." J. Acoust Soc. Am. 20, 259-266 (1948).

[19] W. R. Dyer. "The Masking of Speech by High- and Low-Pass Noise," Tech. Document Rept. No. RADC-TDR-62-298, Rome Air Development Center (1962).

Fig. 2 in Ref. 7 to be equal in over-all level (measured by the C or flat weighting network of an SLM), it is possible to plot them at levels which are equally speech-interfering. This is possible by utilizing the information in Fig. 3 of Ref. 7, which shows the speech-to-(over-all level of) noise ratio for given word scores in the four noises of Fig. 2 (Ref. 7). For an equal word score of, say, 60%, the relative levels among noises can be found by assuming a given level of speech and noting that the over-all level of noise A is 26 dB less than the speech level, noise B is 2 dB less, noise C is 21 dB less, and noise D is 7 dB less. To show the levels at which the noise spectra in Fig. 2 of Ref. 6 are equally speech-interfering (at 60%), the noise spectrum can be re-plotted such that, with respect to noise A, noise B is reduced in over-all level by 24 dB, C by 5 dB, and D by 19 dB. If this is done, the spectrum plots of the equally speech-interfering noises intersect each other between the 600–1200-cps octave and the 1200–2400-cps octave, or at roughly 1200 cps; the steeply sloped spectra below this at, say, 900 cps; and the gently sloped ones above at, say, 1400 cps.

It appears therefore from the evidence of Pollack,[13] Dyer,[29] Kryter,[7] and this study, that noise-masked speech has a crossover or importance frequency as much as an octave lower than the crossover frequencies of filtered speech in the quiet (French and Steinberg,[16] and Beranek[17]). Both Pollack[13] and Dyer[29] show that the frequency varies from 800 or 1000 cps to 1600 or 2000 cps as the speech-to-noise differential increases.

SUMMARY

This paper has examined ways in which 16 noises adjusted in level to be equally speech-interfering can be rated by AI and average curve-fitting methods. It is a continuation of a companion paper[2] in which simpler physical rating or measuring schemes were evaluated.

In this paper it was found that AI procedures are good but only marginally better than the 500-, 1000-, and 2000-cps SIL found best in Ref. 2. It was also observed that the 5- or 6-octave-band procedures are about as good as the 20-band method, and that a generalized speech spectrum was almost as valid as the specific speech spectrum for comparative evaluations among the 16 noises.

The 16 equally speech-interfering noises were neither equally loud nor equally noisy. But their speech-interference value could be predicted very well from families of NC, NCA, ISO (loudness or noys) curves if (1) only that part of any of the curves that center on the octaves at 500, 1000, and 2000 cps is used, and (2) the curves "average through" spectral peaks and valleys. The curve-fitting techniques do not work well if only spectral peaks are allowed to touch them.

Evidence from this and other studies shows that although the "importance frequency" for filtered speech in quiet is around 1700 or 1900 cps this "importance frequency" drops as much as an octave as the noise masking increases, or, more precisely, as the S–N differential decreases from, say, 30 to 5.

PSYCHOPHYSICAL MEASUREMENTS OF EQUALLY SPEECH-INTERFERING NOISES

In Sections III and V of this report, methods of predicting the speech interference properties of 16 diverse noises were discussed. The general procedure was that 16 noises were adjusted in level so that listeners hearing monosyllabic words at a constant level via a loudspeaker obtained 50 percent word intelligibility scores. Then various physical measurements were made on the 16 noises reproduced at the equally speech-interfering levels.

The results showed that, measuring the noise level in octave bands and averaging those bands, the Speech Interference Level concept was the best method if simplicity and accuracy of prediction were both considered. Weighted 6-octave band Articulation Index calculations were slightly better in accuracy of prediction.

This section of the report details other techniques of measurement.

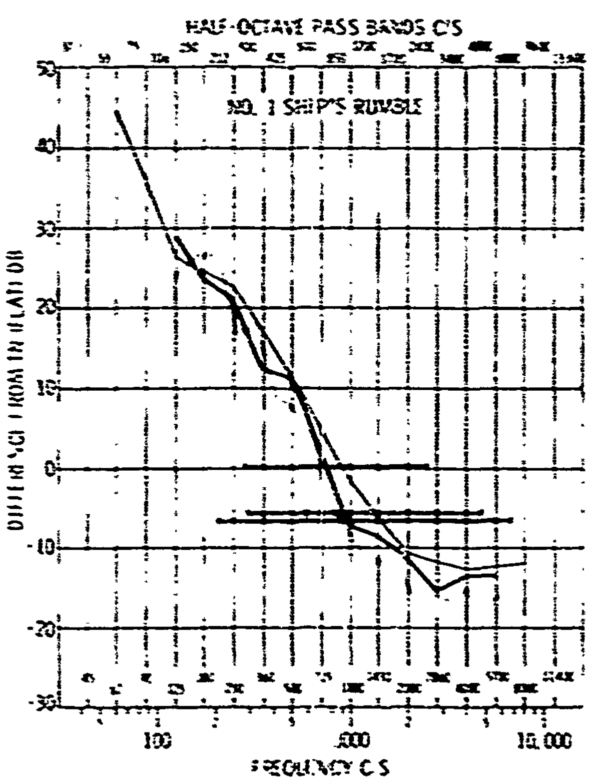
Masked Threshold Spectra

In Sections III and V no physical measurement was presented that was completely satisfactory in specifying that the 16 noises were equally speech-interfering. It was believed that some procedure utilizing the properties of the human ear might lead to better results, so in Section VI some measurements are given that were obtained by utilizing the masking effects of the noises.

Three independent masked audiograms from each of two experienced listeners were found for each of the 16 noises. A Bekesy audiometer, using pulsed, half-octave bands of noise as the probe stimuli, was employed. The half-octave probe stimuli from the Bekesy audiometer and the 16 masking noises were fed into a single Lansing Iconic loudspeaker situated in an acoustically treated room that measured 7 feet by 10 feet by 9 feet. The listener sat one meter away from the loudspeaker and for each of the 16 noises found, by means of an attenuator and a bracketing technique, the level at which each of the half-octave bands of noise centered at 125, 180, 250, 360, ... 4000, and 5700 c/s was just barely audible.

The masking experiments reported in this section were conducted a year after the speech tests described in Section III, IV, and V, and it was not possible to reassemble the same electroacoustic system used in the speech tests, nor to test in the same room. To minimize effects due to the characteristics of the new playback system and test room, all masked-audiogram spectra are plotted as differences from the flat, thermal noise (10). Figures VI-1 to VI-16 (except fig. VI-10) show these masked-audiogram-derived spectra. Also shown in figures VI-1 to VI-9 and VI-11 to VI-16 are the differences from noise 10 of third-octave band levels measured acoustically with the General Radio Sound and Vibration Analyzer, Type 1554A. Likewise, the difference in octave levels of all noises from the noise 10 are shown. The levels on which the octave-level difference are based are taken from Section III. Broadband thresholds (the horizontal bars) are also plotted in figures VI-1 to VI-9 and VI-11 to VI-16, but these will be discussed in Section VII.

——— HALF-OCTAVE MASKED AUDIOGRAM DATA (SHIP'S RUMBLE MINUS THERMAL NOISE)
— —— DIFFERENCE BETWEEN SHIP'S RUMBLE AND THERMAL NOISE MEASURED IN OCTAVE BANDS
——— DIFFERENCE BETWEEN SHIP'S RUMBLE AND THERMAL NOISE MEASURED IN THIRD OCTAVE BANDS
——— DIFFERENCES BETWEEN MASKED THRESHOLDS OF NOISE BANDS FOR SHIP'S RUMBLE AND THERMAL NOISE



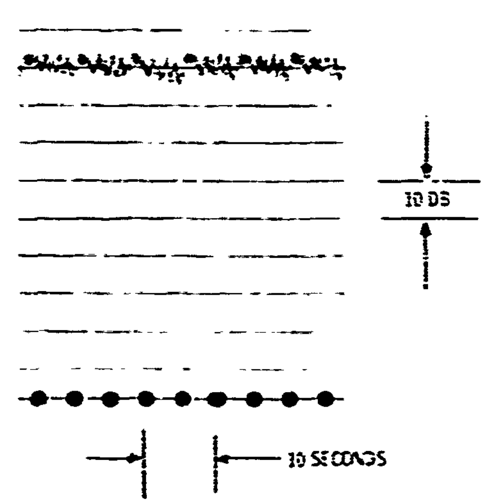
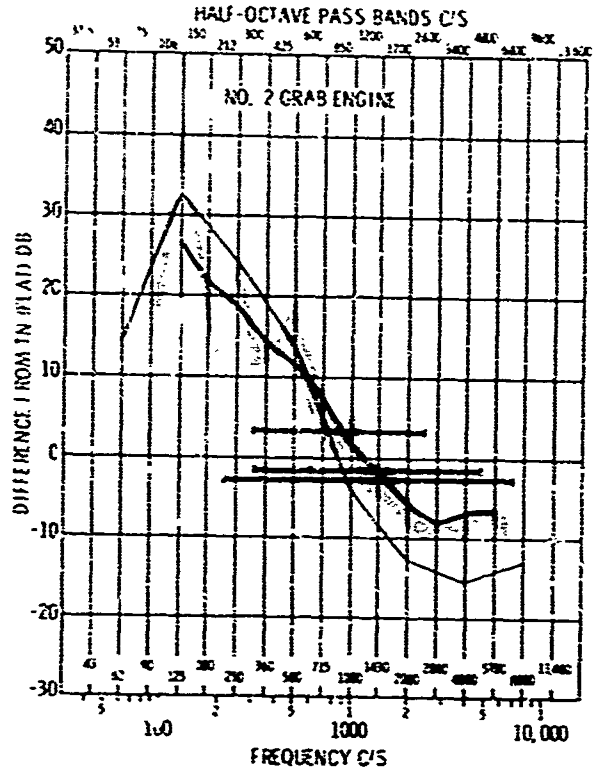


Figure VI-1. Difference from Thermal Noise of Noise 1, Ship's Rumble.



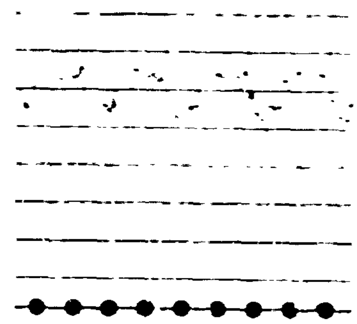
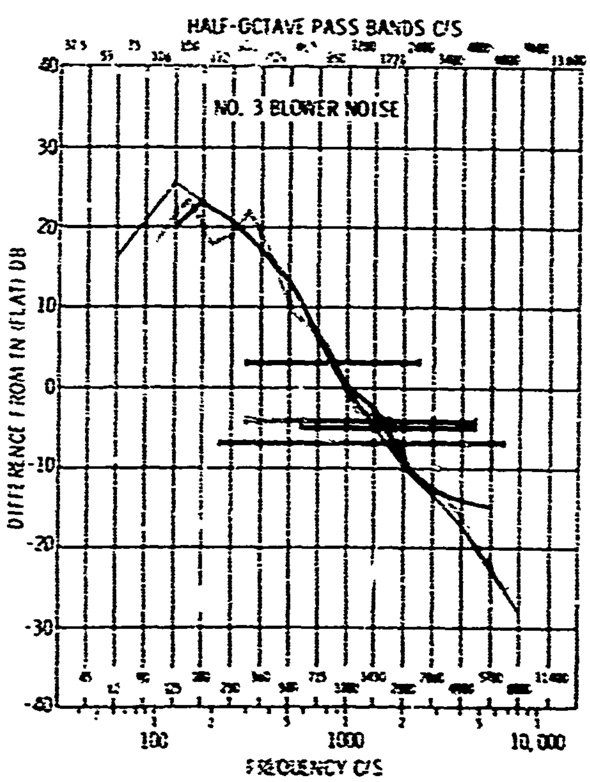


Figure VI-2. Difference from Thermal Noise of Noise 2, Grab Engine. This is a fluctuating noise and so two masked thresholds were found (grey area).



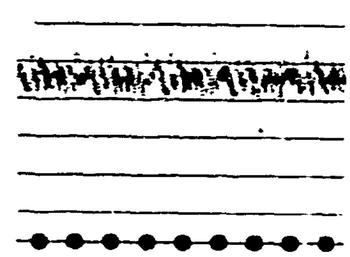
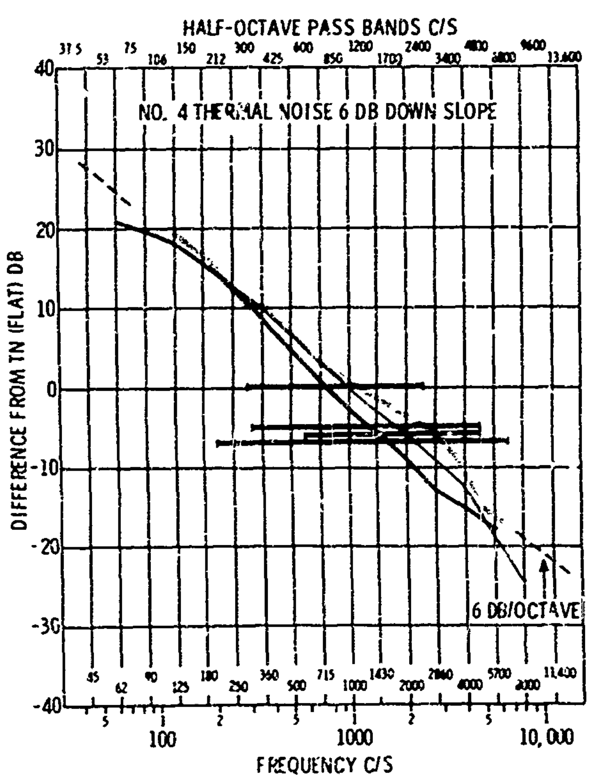


Figure VI-3. Difference from Thermal Noise of Noise 3, Blower Noise.

HALF-OCTAVE MASKED AUDIOGRAM DATA (SHIP'S RUMBLE MINUS THERMAL NOISE)
DIFFERENCE BETWEEN SHIP'S RUMBLE AND THERMAL NOISE MEASURED IN OCTAVE BANDS
DIFFERENCE BETWEEN SHIP'S RUMBLE AND THERMAL NOISE MEASURED IN THIRD OCTAVE BANDS
DIFFERENCES BETWEEN MASKED THRESHOLDS OF NOISE BANDS FOR SHIP'S RUMBLE AND THERMAL NOISE



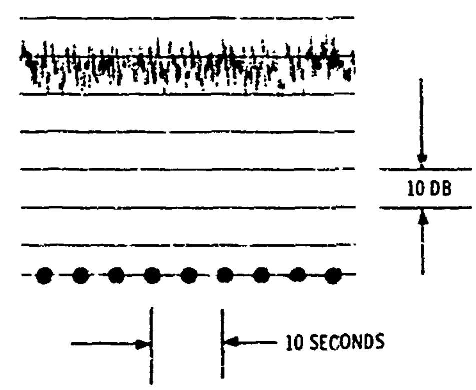
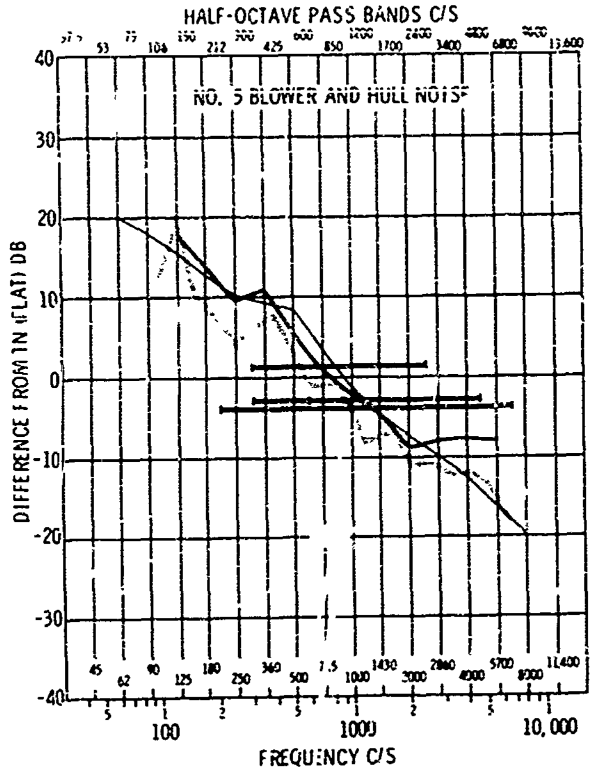


Figure VI-4. Difference from Thermal Noise of Noise 4, Thermal Noise, -6 dB per octave slope.



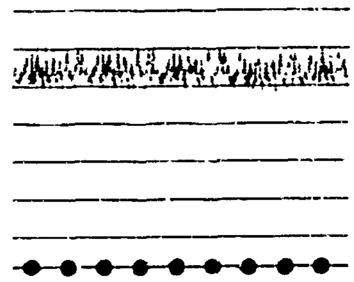
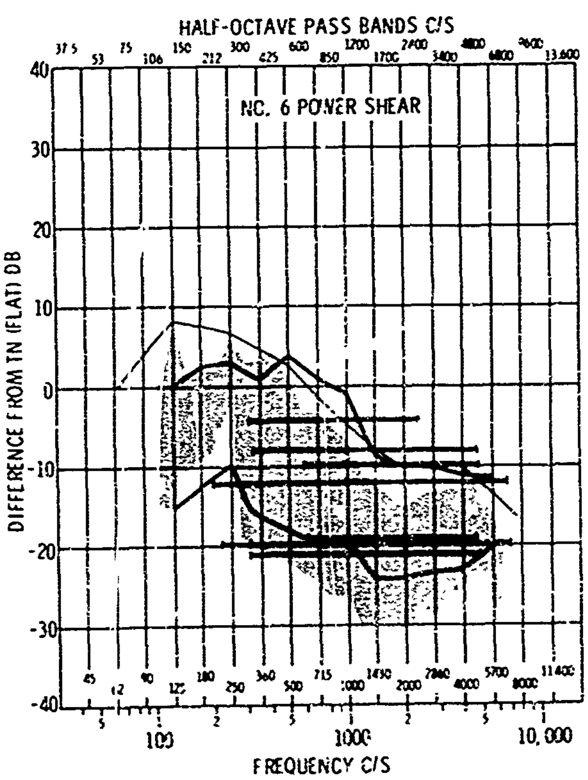


Figure VI-5. Difference from Thermal Noise of Noise 5, Blower and Hull Noise.



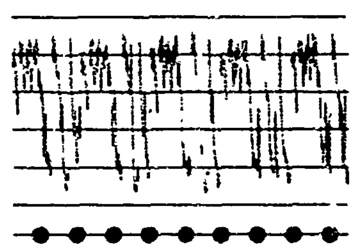
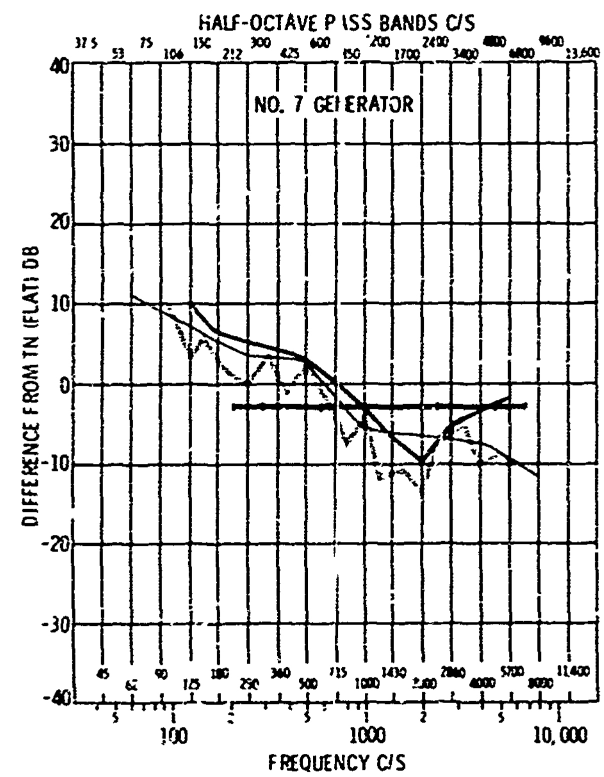


Figure VI-6. Difference from Thermal Noise of Noise 6, Power Shear. This is a fluctuating noise and all masked threshold data (grey area and horizontal bars) have two values.

HALF-OCTAVE MASKED AUDIOGRAM DATA (SHIP'S RUMBLE MINUS THERMAL NOISE)
DIFFERENCE BETWEEN SHIP'S RUMBLE AND THERMAL NOISE MEASURED IN OCTAVE BANDS
DIFFERENCE BETWEEN SHIP'S RUMBLE AND THERMAL NOISE MEASURED IN THIRD OCTAVE BANDS
DIFFERENCES BETWEEN MASKED THRESHOLDS OF NOISE BANDS FOR SHIP'S RUMBLE AND THERMAL NOISE



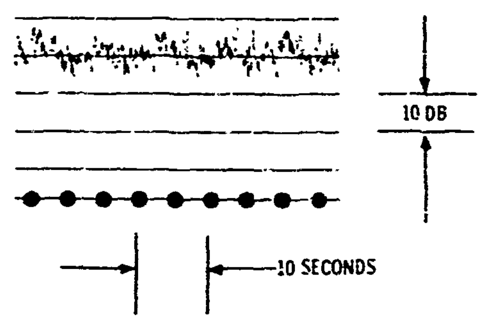
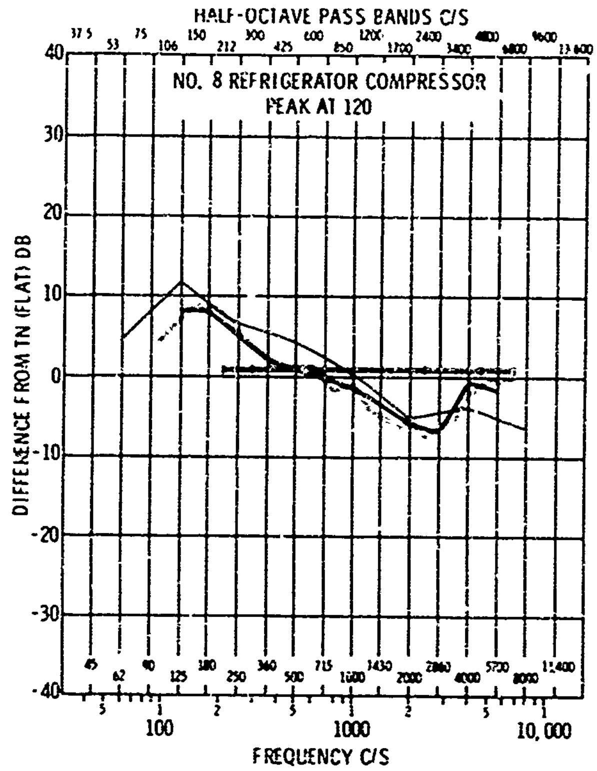


Figure VI-7. Difference from Thermal Noise of Noise 7, Generator.



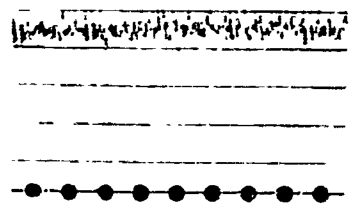
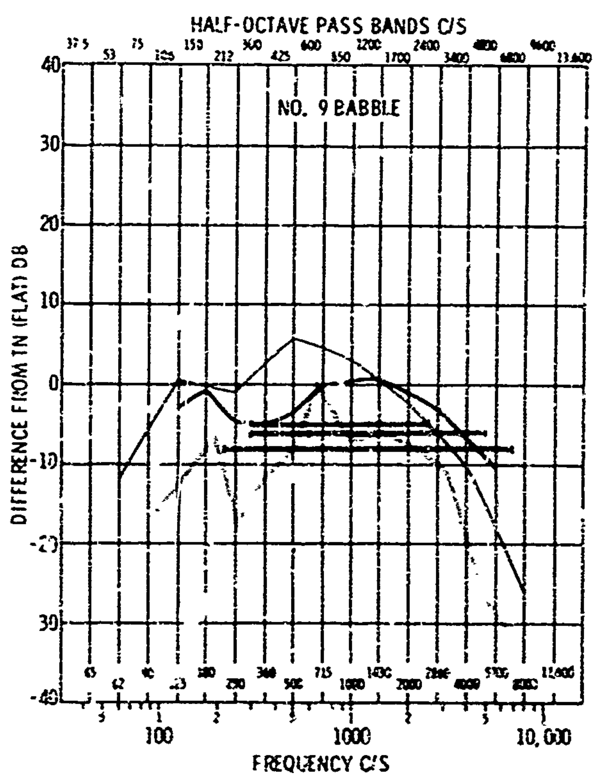


Figure VI-8. Difference from Thermal Noise of Noise 8, Refrigerator Compressor.



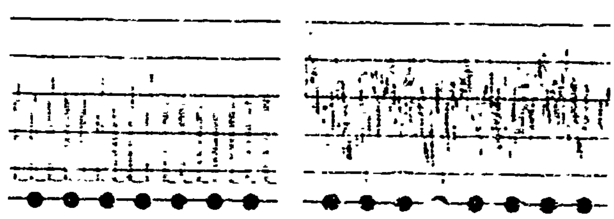


Figure VI-9. Difference from Thermal Noise of Noise 9, Voice Babble. The trace on the left is the level, and timing sequence, of the speech test words; the two peaks are "write, hot," "write, pay," "write, top," etc.

Unlike figures VI-1 to VI-9 and VI-11 to VI-16, figure VI-10 is the actual spectrum of noise 10. The uppermost curve is the octave-band spectrum. It is taken directly from figure 10, Section III, and is therefore representative of the level and the acoustic conditions (including playback system and room acoustics) of the equal-speech-interference part of these studies. The next curve down is the third-octave acoustic spectrum of the thermal noise as used in the masked-audiogram part of these studies. A different room from the previous testing room (smaller and with shorter reverberation time), different playback equipment (Altec Iconic loudspeaker), and a 10-dB lower sound pressure level were utilized. The 10-dB lower level was used because of equipment limitations. The masked threshold of half-octave bands of noise heard in the presence of the level and spectrum of thermal noise depicted by the third-octave analysis is shown by the solid squares.

These three spectra - the octave, third octave, and masked threshold levels - are those used for the zero or reference lines on figures VI-1 to VI-9 and VI-11 to VI-16, that is, on these figures, when octave level differences are plotted, the octave level shown in figure VI-10 is the common spectrum or zero reference line. When third-octave difference spectra are depicted in figures VI-1 to VI-9 and VI-11 to VI-16, the third octave spectrum on figure VI-10 is the common or referral spectrum. When half-octave masked-threshold differences are plotted, the zero reference is the masked threshold data plotted in figure VI-10.

The reason for utilizing the difference spectra in figures VI-1 to VI-9 and VI-11 to VI-16 was to equate out of each of the 16 noises the amplitude vs frequency response of the playback system (recorder, loudspeaker, and room), and thereby facilitate comparisons with the earlier studies. As is evident in figure VI-10, the playback system had a generally rising characteristic to 360 c/s, a dip at 500 c/s, and a falling characteristic above 2500 c/s. But all 16 noises had this same relative response; therefore the use of difference spectra essentially eliminates this as a source of undue complexity.

The bottom of figure VI-10 shows the spectrum level of the thermal noise, based on the third-octave spectrum data. The difference between the masked threshold data and the spectrum level data increases with rise in frequency, i.e., from 20 to 30 dB between 250 to 5000 cycles. This difference, according to Hawkins and Stevens,[16] is sometimes taken as defining...the critical bandwidth of a masking noise." They were concerned with pure tones masked by white noise, whereas the present data concern half-octave bands of noise masked by white noise. Greenwood,[17] in an extensive study of critical bandwidth, used narrow bands of noise as probes, but his noises were much narrower than half-octaves and his masker was pure tones, not broadband noise. Greenwood's data on pure tones agree, however, with Hawkins and Stevens. Perhaps the similarity (of noise in noise) vs the distinctness (of tones in noise), accounts for the 5-dB difference noted between the two curves shown at the bottom of figure VI-10. However, differences in mode of presentation (earphone vs loudspeaker), choice of method (adjustment vs Bekesy audiometer), monaural vs binaural listening, or between Hawkins and Stevens, the two subjects, and those of Greenwood, could contribute to the difference.

OCTAVE BAND LEVEL FROM DATA IN SECTION III
MASKED THRESHOLD MEASURED IN HALF OCTAVES
THIRD-OCTAVE LEVEL
SPECTRUM LEVEL
BANDWIDTH (IN DB) OF HALF-OCTAVE BANDS OF NOISE IN THERMAL NOISE (MASKED THRESHOLD DATA MINUS SPECTRUM LEVEL)
CRITICAL BANDWIDTH FOR PURE TONES IN NOISE

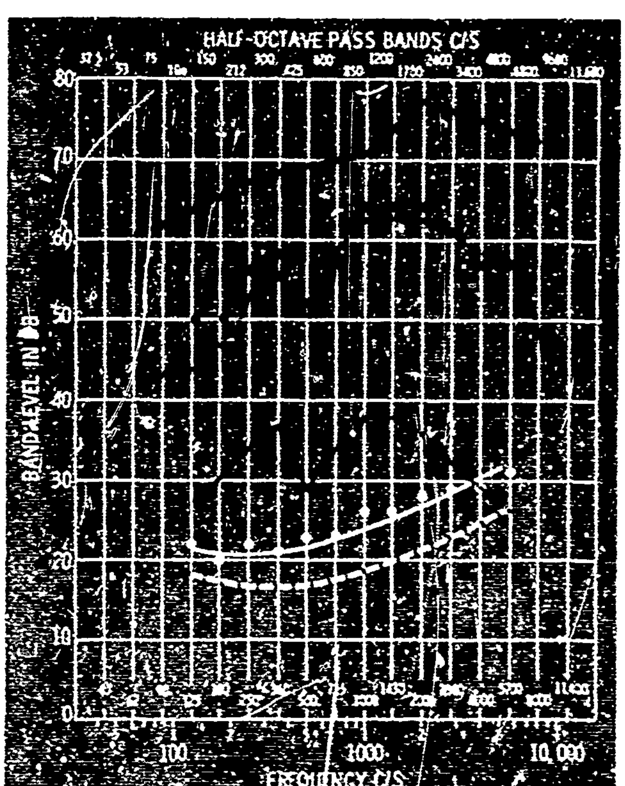
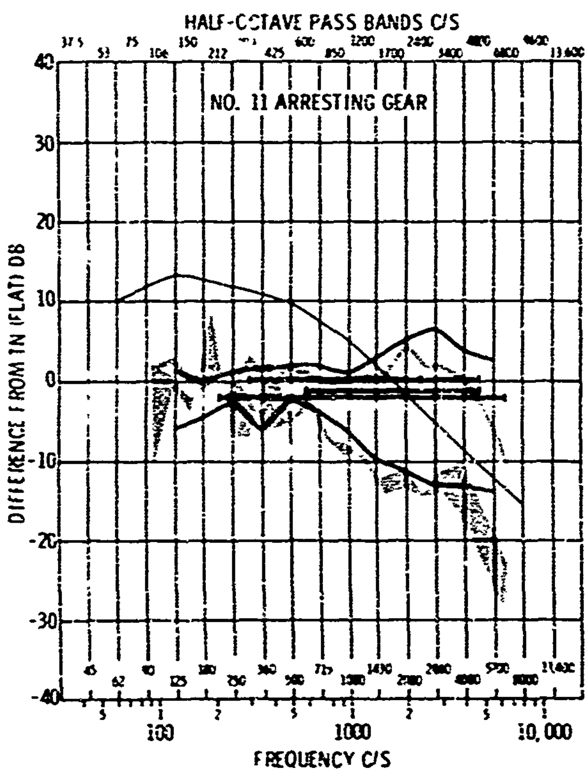


Figure VI-10. System response of thermal noise into loudspeaker in sound-treated, but non-anechoic room. This is the reference noise to which all other noises (fig. VI-1 to 9 and 11 to 16) are compared. No log amplitude-time flat is shown but the time variation is just slightly greater than the trace shown on figure VI-15. The upper three black curves are physical measures. The spectrum level is calculated from the third-octave data.

HALF-OCTAVE MASKED AUDIOGRAM DATA (SHIP'S RUMBLE MINUS THERMAL NOISE)
DIFFERENCE BETWEEN SHIP'S RUMBLE AND THERMAL NOISE MEASURED IN OCTAVE BANDS
DIFFERENCE BETWEEN SHIP'S RUMBLE AND THERMAL NOISE MEASURED IN THIRD OCTAVE BANDS
DIFFERENCES BETWEEN MASKED THRESHOLDS OF NOISE BANDS FOR SHIP'S RUMBLE AND THERMAL NOISE



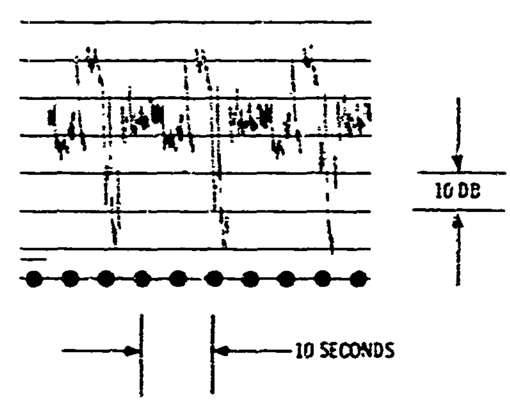
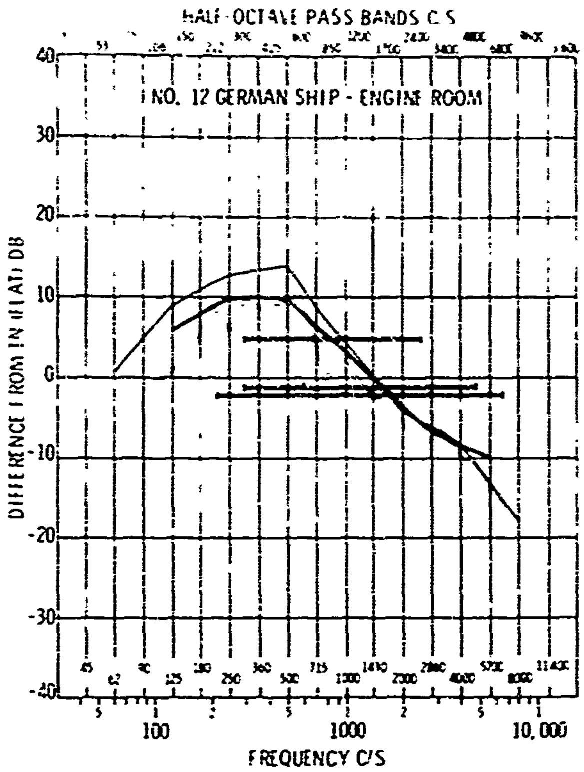


Figure VI-11. Difference from Thermal Noise of Noise 11, Arresting Gear. There are three masked threshold spectra, one for each of three levels of fluctuating noise.



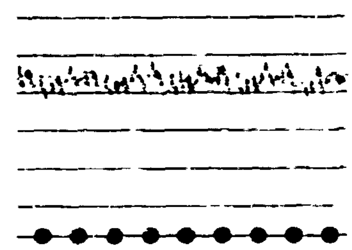
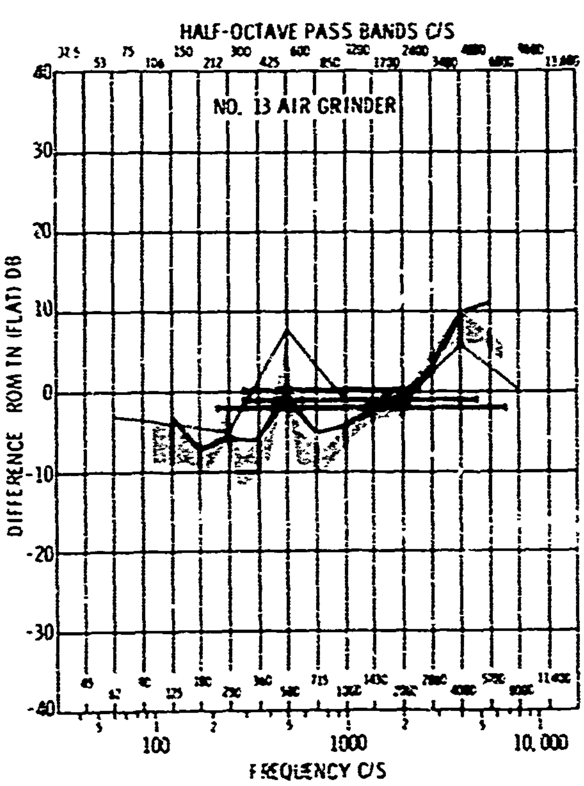


Figure VI-12. Difference from Thermal Noise of Noise 12, German Ship Engine Room.



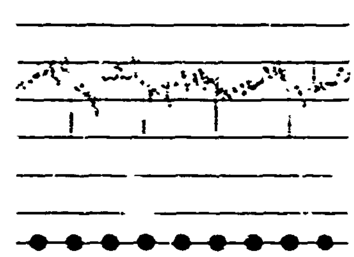


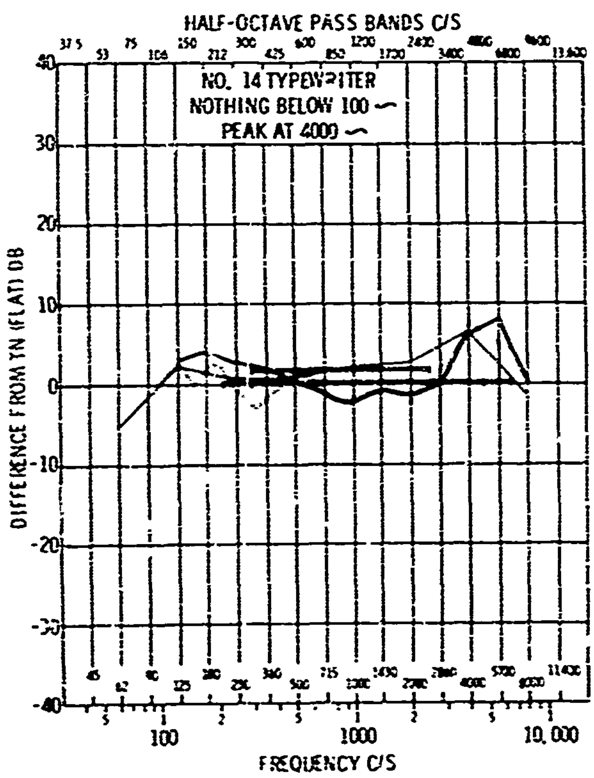
Figure VI-13. Difference from Thermal Noise of Noise 13, Air Grinder. This is a fluctuating noise and two masked threshold (grey area) are shown.

HALF-OCTAVE MASKED AUDIOGRAM DATA (SHIP'S RUMBLE MINUS THERMAL NOISE)

DIFFERENCE BETWEEN SHIP'S RUMBLE AND THERMAL NOISE MEASURED IN OCTAVE BANDS

DIFFERENCE BETWEEN SHIP'S RUMBLE AND THERMAL NOISE MEASURED IN THIRD OCTAVE BANDS

DIFFERENCES BETWEEN MASKED THRESHOLDS OF NOISE BANDS FOR SHIP'S RUMBLE AND THERMAL NOISE



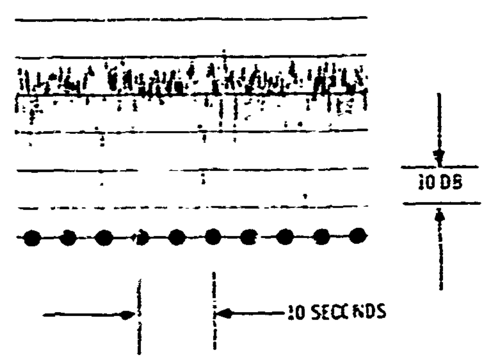
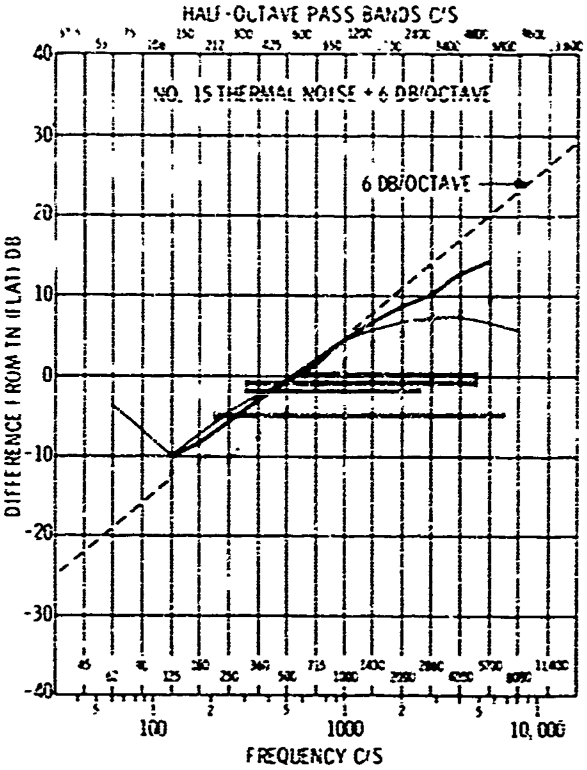


Figure VI-14. Difference from Thermal Noise of Noise 14, typewriter.



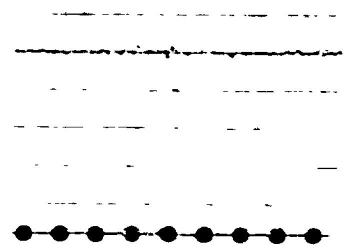
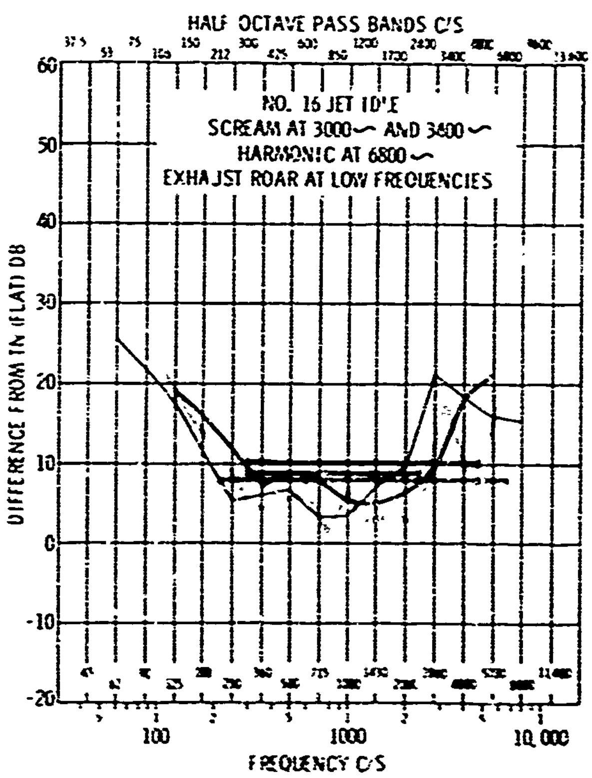


Figure VI-15. Difference from Thermal Noise of Noise 15, Thermal Noise, +6 dB per octave slope.



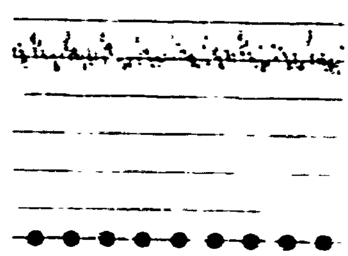


Figure VI-16. Difference from Thermal Noise of Noise 16, Jet Idle.

Figures VI-1 to VI-16 show that in general the half-octave masking data (heavy solid line) are closely equivalent to the third-octave filter data (grey line). There is not a great deal of upward masking for the noises with most of the energy at low frequency shown in figures VI-1, VI-2, VI-3, and VI-4. That is, the heavy solid line in these four figures does not lie consistently above (nor to the right of) the grey line. In figures VI-5, VI-6, and VI-7, there may be some upward masking. Figure VI-15 shows the reverse, a downward shift. The octave band data do not always agree too well with either the third-octave physical data or the half-octave, masked-audiogram data. It should be pointed out that the octave results are derived from earlier data (Section III) where a different audio system was used for reproducing the noises. The largest discrepancies occur on the most fluctuating noises, Nos. 6, 9, 11, and 14.

Note that the half-octave masked threshold data often level out the peaked spectra information shown in the data derived from the third-octave filters (see Nos. 2, 6, 9, 11, 13, and 16). This is because the third-octave filters "see" everything within their passband, and peaks (tonal components) result in high numerical readings. When the listeners find the threshold for a half-octave band of noise in the presence of these same peaks, they essentially hear around the tonal components and base their threshold on parts of the banded noise outside or beyond the peaked or tonal components.

Broadband Masked Thresholds

Also shown in figures VI-1 to VI-16 (except fig. VI-10) are broadband-noise thresholds for each noise. Four broadband noises are used: 300 to 2400-, 300 to 4800-, 600 to 4800-, and 200 to 6800-c/s bands. As on all plots in figures VI-1 to VI-9 and VI-11 to VI-16, these broadband thresholds are expressed as differences from the respective broadband thresholds found for the thermal noise. Although not plotted in figure VI-10, the broadband masked thresholds in the presence of 72 dB of thermal noise were 65 dB for the 300 to 2400-c/s band; 67 dB for the 300 to 4800- or 600 to 4800-c/s band; and 68 dB for the 200 to 6800 band.

In figures VI-1 to VI-9 and VI-11 to VI-16 the broadband thresholds, expressed as differences, are indicated by horizontal lines that define the bandwidth of the band. As would be expected, the threshold for these broad-bands of noise in the presence of predominantly low-frequency noises is determined by the high-frequency cutoff of the probe band noise. The threshold is successively lower as the high-frequency cutoff of the probe noise band is successively higher (noises and figures VI-1 to VI-6 and to a lesser extent, noises and figures VI-9 and VI-12). For the flatter spectrum noises (Nos. 7, 8, 11, 13, and 14), the broadband thresholds are roughly equivalent regardless of either high-or low-frequency cutoff. For the one clear-cut, high-frequency noise (No. 15), the noise bands with the lowest frequency cutoffs have the lowest thresholds. Note that a line drawn through the upper cutoff frequencies of the probe noise bands is roughly parallel to the masking noise spectra on the low-frequency noises. And conversely, a line through the lower cutoff frequencies of the pulse noise bands is parallel to the high-frequency masking noise spectrum in figure VI-15.

The question is posed as to whether the masked thresholds of these broadband probe noises, which were chosen to lie in the important speech frequency regions, will give a measure of the speech-interfering properties of the 16 masking noises. It is important to know how much variation there is in the 300 to 2400-c/s band threshold among the 16 noises. In columns 1 and 2 of table VI-1 are listed the number and names of the 16 noises. Column 3 is a weight that reflects an estimate of how often noises of this type would occur in larger samples of ship and possibly industrial noises. Column 4 lists the 300 to 2400-c/s band thresholds as read directly from figures VI-1 to VI-16 (No. 10 is by definition zero since every noise is compared to it). The mean and standard deviation of these 16 numbers, the mean and standard deviation of the 33 numbers represented by three noises like No. 1, one like No. 2, five like No. 5, etc.), and the extent of the spread between the highest and lowest number (the range) are listed below columns 4 through 12. Columns 5, 6, and 7 list the other broadband-noise thresholds taken from figures VI-1 to VI-16. Column 8 lists the threshold obtained using a voice babble as the probe stimulus to assess the masking of the 16 noises. These voice babble data are not plotted in figures VI-1 to VI-16.

TABLE VI-1. PHYSICAL AND PSYCHOPHYSICAL MEASURES OR CALCULATIONS

Column 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
			Broadband Thresholds					Physical SIL's					Psychophysical SIL's	
	Noise	Weight	3-24	3-48	6-48	2-68	Babble	3-band 500-2K	4-band 250-2K	600-4800	300-4800	300-2400	3-band 500-2K	4-band 250-2K
1.	Ship Rumble	3	0	-6	-6	-7	-7	71.6	76.3	66.3	70.5	73.0	-5.3	0.1
2.	Grab Eng.	1	3	-2	-2	-3	-5	71.1	76.3	64.4	69.0	71.7	0.5	4.4
3.	Blower	3	3	-4	-5	-7	-5	73.7	77.3	67.0	71.7	75.4	-1.0	3.9
4.	TN-6	2	0	-5	-6	-7	-5	73.1	75.1	69.3	71.7	73.7	-4.2	-0.3
5.	Blower and Hull	5	1	-3	-3	-4	-6	72.3	73.7	68.6	71.4	73.2	-3.2	0.3
6.	Shear	1	-4	-8	-10	-12	-10	68.5	70.0	68.9	70.0	69.2	-4.2	-2.3
7.	Generator	2	-3	-3	-3	-3	8	69.6	70.0	69.9	70.7	70.1	-3.3	-1.4
8.	Compressor	3	1	0	1	1	7	72.0	72.6	71.4	72.1	72.4	-3.2	-1.5
9.	Babble	3	-5	-6	-5	-8	14	74.8	72.8	73.2	73.2	74.4	-1.0	-2.0
10.	TN Flat	1	0	0	0	0	0	72.6	71.4	74.6	73.4	72.3	0.0	0.0
11.	Arresting Gear	1	0	0	-1	-2	0	77.3	77.9	75.2	76.8	78.1	4.0	3.4
12.	Engine Rm.	2	5	-1	-1	-2	4	77.4	78.2	74.2	76.7	78.9	1.7	3.8
13.	Air Grinder	1	0	-1	-1	-2	0	75.3	72.0	75.7	76.0	74.2	-1.7	-2.8
14.	Typewriter	3	2	0	-1	0	0	74.8	73.3	78.0	76.3	74.1	-1.0	-0.1
15.	TN+6	1	-2	-1	0	-5	2	75.9	72.7	80.0	77.0	74.8	4.8	2.5
16.	Jet	1	9	10	8	8	9	79.0	77.4	81.0	79.8	[illegible]	6.8	7.5
	Range		14	18	18	20	24	10.5	7.9	16.6	10.8	9.7	12.1	10.3
	Mean (16)		0.6	-1.9	-2.2	-3.3	0.4	73.7	74.2	72.4	73.0	73.9	-0.6	1.0
	Stand. Dev. (16)		[illegible]	3.9	3.9	4.5	6.5	2.8	2.7	4.8	3.1	2.5	3.4	[illegible]
	Stand. Dev. (33)		3.0	3.2	[illegible]	3.8	6.8	2.4	2.4	4.3	2.6	2.1	2.8	2.4
	Mean (33)		0.3	2.5	2.7	3.7	0.2	73.4	74.2	71.4	73.0	73.8	-1.6	0.6

Psychophysical Speech Interference Levels

In Section III the concept of the SIL was explained and calculations using a variety of different combinations of octave-bands were tabulated (table II Section III). To show comparisons between the masked threshold data and some of the better methods in Section III, five columns of SIL calculations are reproduced from Section III and listed in columns 9 through 13 of table VI-1.

In columns 14 and 15 are similar SIL scores based on the spectra of figures VI-1 through VI-16. These are called psychophysical SIL's because they are calculated from spectra arrived at by the masked threshold technique. Since all noise spectra in figures VI-1 through VI-9 and VI-11 through VI-16 are difference spectra (the reference or zero line represents the spectrum of the flat, thermal noise), the SIL's are all relative to an SIL of zero for the flat, thermal noise), the SIL's are all relative to an SIL of zero for the flat, thermal noise (No. 10). Columns 14 and 15 show the psychophysical SIL's for both the three- and four-band set of octaves based on the "preferred" center frequencies of 250, 500, 1000, and 2000 c/s.

To determine which of columns 4 through 15 in table VI-1 has the smallest dispersion among the 16 noises, the standard deviations must be examined. Column 13 has the smallest standard deviation. However, if the standard error of the standard deviation ($1/\sqrt{2N}$) is considered, columns 9 through 14, except for column 11, must be adjudged equivalent. Columns 4, 5, and 6 could hardly be excluded statistically, and it could be considered that columns 9 to 14 (except 11) are in one class, columns 4 to 7 and 11 in another, and column 8 off by itself. In any grouping it is evident that the physical SIL's from Section III are not statistically different from the psychophysical SIL's calculated from the data in this section.

Figure VI-17 summarizes the SIL and broadband threshold data tabulated in table I (Section III). In every case, the measures have been set equal on noise 10. From the present data, and those in Section III, it would appear that any SIL taking into account the octave below 600 c/s would predict the speech-interfering properties of these noises better than the presently used 600 to 4800-c/s SIL. The psychophysical SIL's are good predictors, but not superior to physical SIL's and much more difficult to determine. Data derived by aural detection of selected noise bands in wider noises are thus neither or greater nor less value than those from SIL methods.

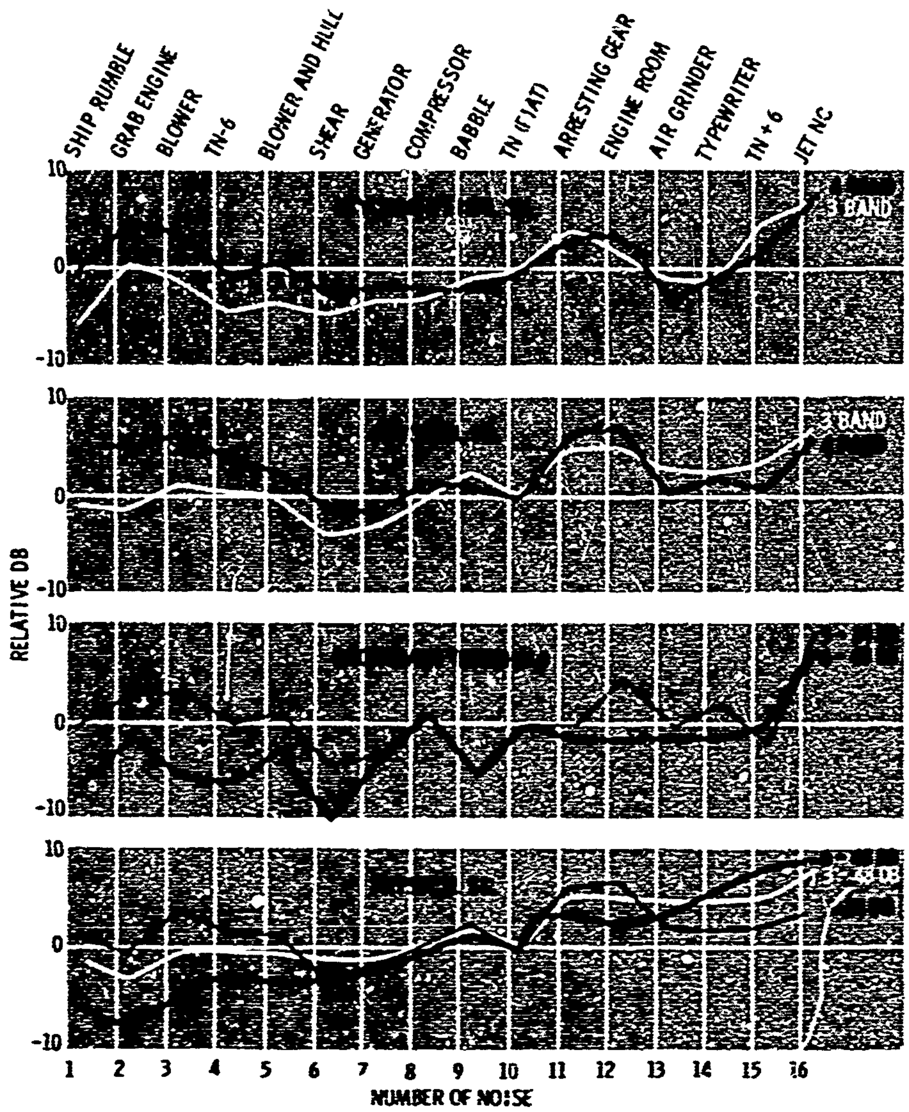


Figure VI-17. Speech Interference Levels (SIL's) for 16 equally-speech-interfering noises. All levels are relative since they are equated on Noise 10, Thermal Noise. The psychophysical and physical SIL's at the top are averages of levels in the octave bands centered at 250, 500, 1000, and 2000 c/s (four-band) or 500, 1000, and 2000 c/s (three-band). The physical SIL's at the bottom are based on averages of levels in the octave bands between 300 and 2400 c/s, 300 and 4800 c/s, and 600 and 4800 c/s. The psychophysical SIL's (at the very top) are based on the half-octave, masked-threshold, difference spectra plotted in figures VI-1 to VI-16. The physical SIL's are from the octave levels printed numerically on figures 1 to 16 of Section III (and equated to Noise 10). The broadband threshold data are from the horizontal-bar results in figures VI-1 to VI-16.

A SPEECH INTERFERENCE NOISE RATING CONTOUR

ABSTRACT

An attempt has been made to show the similarities between three ostensibly different methods of rating noises for speech interference. The three basic methods are: Sound Level Meter (SLM) readings using various frequency weighting networks; Noise Criteria Contours, where spectral peaks of noise become tangent to one of a family of rating curves; and average-level methods, the Articulation Index (AI) being the most sophisticated method and the Speech Interference Level (SIL) being the simplest to use.

A Speech Interference (SI) curve has been evolved which, when used as a frequency weighting network in a SLM, or as a noise-rating curve, or as a curve-fitting method of arriving at an SIL, greatly reduces the spread of scores among the three measurement methods when rating the speech-interfering properties of certain 16 noises.

INTRODUCTION

In a recent series of papers by Klumpp and Webster (1, 2) and Webster and Klumpp (3), physical and psychophysical schemes were examined that purported to measure the speech-interfering aspects of noise. Sixteen diverse-spectrum noises were adjusted in level so that listeners hearing monosyllabic (Rhyme) words at a constant level of 78 dB from a loudspeaker obtained 50 percent word intelligibility scores. Twenty-band, and 5- or 6-octave-band Articulation Index (AI) calculations, see Kryter (4), predicted the speech-interfering properties of the noises very well, see Webster and Klumpp (3). However, as shown by Klumpp and Webster (2), some other, and simpler, schemes worked just as well; for example, Speech Interference Level (SIL) calculations, see Beranek (5), based on octaves centered at 425, 850, and 1700 c/s, or 500, 1000, and 2000 c/s. The A-weighting and Din 3 networks, see Peterson and Bruel (6), of a Sound Level Meter (SLM) were good, but the conventional use of Noise Criteria (NC), or Alternate Noise Criteria (NCA), see Beranek (7), curves did not work well. However, NC, NCA, and ISO, see Janssen (8), curves worked very well if (1) only that part of the curves centering on the octaves 500, 1000, and 2000 c/s was used, and (2) the noise spectra were allowed to "average through" a contour and not just touch it at a peak value. This "average through" or average-curve-fitting method is a combination of methods. It uses contours customarily used in the tangent-to-curve method to arrive at the equivalent of an SIL (average-level method). In the discussions which follow it will not be spelled out as a separate method but will be considered as just another average-level method. See ref. (3) for more discussion on this combination method.

In the process of trying all possible noise-rating schemes, it became evident that there were essentially three basic ways to rate the speech interference properties of noises. And although the three basic methods differ in how they operate, the best of each method was pretty good and with a few compromises here and there the three basic simple methods might become quite comparable.

METHODS OF RATING SPEECH INTERFERENCE

The three basic methods of rating the speech-interfering properties of noise are: (1) average-level methods (the AI being the most comprehensive, universal, and the best predictor, but the SIL doing as good a job if the proper octaves are chosen initially); (2) SLM frequency weighting networks (A and Din 3 being conclusively better than either B or C); and (3) tangent-to-curve methods.

As stated in Klumpp and Webster (2), these methods work in different ways and it is pertinent to point out how they differ. The simplest in concept, but the worst in predictive ability, is the tangent-to-curve method. In this method, only the noise component (peak) that first touches a generalized noise-rating contour determines the rating. Any pure tone component, or any restricted band component, that differs drastically from its surroundings dominates this rating.

The tangent-to-curve method may be expressed mathematically as follows:

$$NR = \text{ten} \log_{10} k(f,p) p_m^2 \; p_o^2 \qquad (1)$$

where NR is any noise-rating criteria desired such as NC, NCA, or ISO; and k is a frequency- and sound-pressure-dependent weighting factor (represented by families of NC, NCA, or ISO contours); p_m is the maximum sound pressure (the noise spectral peak that first touches a given contour); and p_o is a reference sound pressure (usually 0.0002 microbar). The magnitude of NR is the logarithm of the weighting factor at the frequency of p_m and the maximum noise sound pressure.

For the weighted-integration (SLM or network) method, the indicating instrument following the weighting network in a SLM adds components powerwise, i.e. two equal components result in a level 3 dB greater than either level individually; nevertheless, as in the tangent-to-curve method, a single component 10 dB greater than all its neighbors essentially determines the level. The frequency weighting network method can be expressed mathematically as follows:

$$w = 10 \log_{10} \left| \sum \left[k_1(f,p)p_1^2 + k_2(f,p)p_2^2 + \ldots k_n(f,p)p_n^2 \right] \; p_o^2 \right| \qquad (2)$$

where w is a sound-level weighting reading; k_1 is a frequency-dependent weighting factor determined by the definition of w; $p_1 \ldots p_n$ are sound pressures in contiguous bands.

The magnitude of w is the logarithm of weighted sums of squared band sound pressure.

The average-level methods (AI and SIL) work conversely. Whereas the tangent and network methods are determined by one (tangent) or more (network) sound pressure peaks and give readings equal to (tangent) or greater than (network) the highest peak sound pressure level, the average-level methods yield measures lower than any single peak by the inclusion of lower levels.

The average-level method can be expressed mathematically as follows:

$$\mathrm{SIL}_{1-n} = 10 \log_{10} \left\{ \left[k_1(f,p) p_1^2 \times k_2(f,p) p_2^2 \times \ldots k_n(f,p) p_n^2 \right]^{1/n} \quad p_0^2 \right\} \tag{3}$$

where SIL_{1-n} is a Speech Interference Level of n bands, both the number of, and the location of the bands must be specified; k is a frequency and sound pressure level dependent weighting factor (but for SIL calculations in the past, k_1 has been equal to $k_2 \ldots k_n = 1$); $p_1 \ldots p_n$ are sound pressure levels in specified bands. As a consequence of the properties of logarithms an equivalent SIL result can be obtained by taking the arithmetic mean of the levels and adding a constant which will depend on the weighting -- a process less involved than taking the individual differences and averaging, but a process that can be used only if k is neither zero nor a function of p; i.e. all SIL differ by constants. The magnitude of an SIL is the logarithm of a product of weighting factors plus the logarithm of a harmonic mean of band sound pressure.

As an example of how these three methods differ, consider a noise that had equal sound pressure level, say, 80 dB, in each of four pertinent octave bands. This noise would be rated 80 by the average-level method, 86 by the integration method, and 80 by the tangent-to-curve method. A tonal component of 90 dB in one band would change the average level to 82.5, change the network reading to 91.1 and the tangent-to-curve rating to 90. Two 90-dB tonal components in adjacent octaves would change the average level to 85, the integration-method reading to 93.4 and the tangent rating to 90. Two 90-dB tonal components in the same band would change the average-level to 83.3, the integration-level to 93.6, and the tangent rating to 93.2. Licklider and Guttman (9) have shown that tonal components do not mask speech very effectively so the speech interference properties of these four hypothetical noises would be approximately equal.

To summarize the example: The average-level measures on these hypothetical noises varied from 80 to 85, the integrated levels from 86 to 93.6, and the tangent rating from 80 to 93. The weighted 5-octave band AI would act much the same as the average-level measure except that weighted averages would be involved and, in fact, in the examples above, changes in AI, expressed in dB, could be as large as 5.9, depending on which octaves the two tones were in and whether the level of speech kept the speech-to-noise (S-N) within the 0 to 30 dB range. To express AI in dB, recall that an AI of 0 corresponds to an average S-N differential of 0, an AI of 1 corresponds to a 30 dB S-N, an AI of 0.5 to 15 dB, etc., so any value of AI can be expressed as some S-N between 0 and 30.

It follows from the above discussion and from the results presented in Klumpp and Webster (2) that on any given noise the integration methods will give the highest numerical ratings (two equal peaks together add 3 dB), the tangent-to-curve method next (highest peak, or peaks, determines rating, no summing), and the averaging methods the lowest ratings. This is strictly true only if the frequency-weighting networks have the same general frequency vs. level shape as the inverse of the tangent-to-curve rating contours.

Figure 1 shows the general similarities among the common SLM weighting networks and the common families of noise-rating contours. Note, for example, that the loudness, see Stevens (10), annoyance, see Kryter (11) and Kryter and Pearsons (12), and NC curves are quite similar in shape, especially at the low frequencies and for the "70" contours, and that all of them tolerate less low-frequency sound than the NCA curve. It should be pointed out that loudness and annoyance calculations are not just simple tangent-to-curve calculations. They are different inasmuch as components other than the peak spectral component are taken into account. In this regard they act more like SLM weighting networks. Also note that for low-frequency sounds at least, the NCA-70 rating curve and the A weighting network curve are very similar in shape. Young (13) has already made a case for using the A weighting network for rating sounds because of the general similarities in shape between the A and the NC-40 contours. Young (14) has now shown that even loudness and annoyance, at least for office noises, are fairly well predicted by the A weighting network.

Since the A network is very similar in shape to the inverse of the NC and/or NCA contours, it is not surprising that the ratings assigned to the 16 equally speech-interfering noises in table 2 of Klumpp and Webster (2) are higher in magnitude on A weighting (83.5 dB) than on NCA curve-limiting (78.7 dB), and that both are larger than the SIL (73.7 dB). (These data are reproduced in columns 5, 8, and 13 of table 1.)

The absolute magnitude of the ratings assigned by variants of the three basic methods is not, however, the

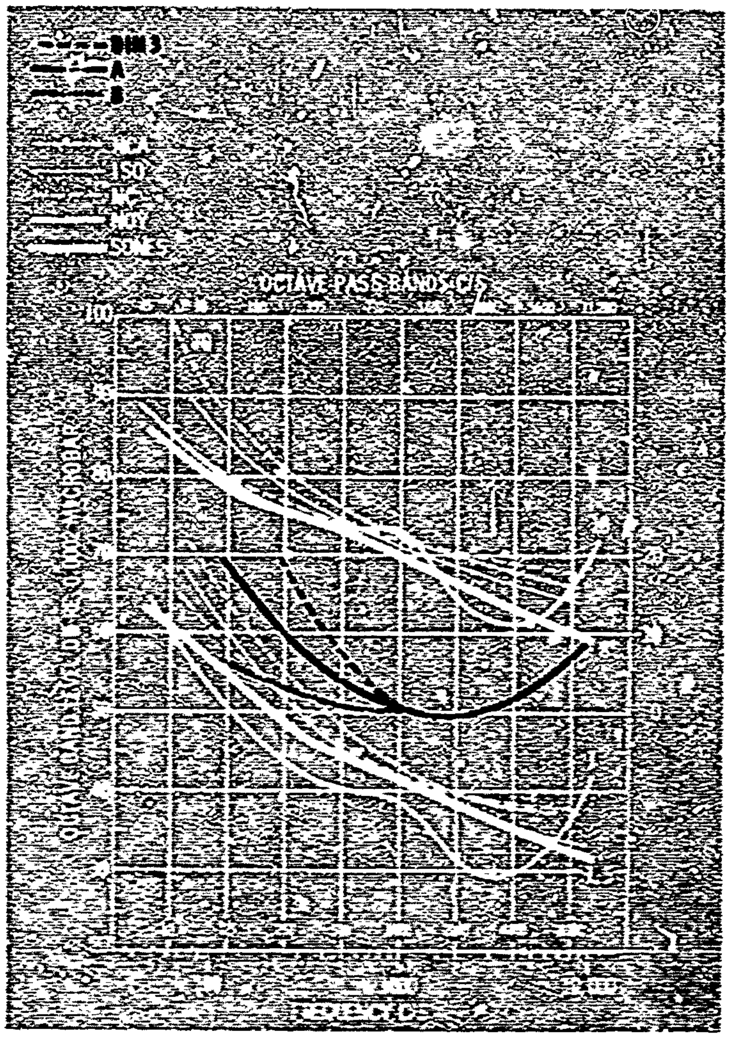


Figure VII-1. Noise rating curves and sound level meter frequency weighting networks.

most important facet. It is the dispersion of the ratings assigned to the 16 equally speech-interfering noises that is important. In this regard, according to Klumpp and Webster (2) and Webster and Klumpp (3), the averaging methods are generally superior (the AI, SIL, or average curve-fitting), the frequency-weighting networks next, and the tangent-to-curve methods worst. (This can be observed by noting that both the Ranges and Standard Deviations increase from column 13 to column 5 to column 8 of table 1.)

TABLE VII-1. VARIABILITY ASSOCIATED WITH
VARIOUS MEASUREMENT PROCEDURES

1	2	3	4	5	6	7	8	9	10	11	12	13	14
		Weighting Networks					Tangent-to-Curve				Averaging or SIL's		
	Noise	C	C(R)	A	A' 70	A'(R) 70	NCA 70	ISO(R)	NCA' 70	Phons	A' 70	3 Band	A'(R) 70
1.	Ship Rumble	105.3	80	86.3	79	76	80	77	75	100	64	71.6	68
2.	Grab. Eng.	97.6	83	86.3	81	79	81	80	78	97	63	71.1	68
3.	Blower	92.3	84	85.1	81	79	79	82	78	94	64	73.7	70
4.	TN-6	86.8	78	79.3	77	76	73	74	72	89	62	73.1	70
5	Blower & Hull	85.3	79	78.8	76	75	73	75	73	89	56	72.3	69
6.	Shear	78.8	74	75.8	73	72	69	71	67	86	59	68.5	66
7.	Generator	79.3	75	78.6	72	72	72	74	68	88	60	69.6	66
8.	Compressor	81.0	77	78.5	74	74	73	74	71	90	62	72.0	69
9.	Babble	80.3	79	80.8	77	77	76	77	74	87	61	74.8	71
10.	TN Flat	80.7	80	81.6	76	76	78	80	72	94	62	72.6	69
11.	Arresting Gear	85.3	82	83.8	80	80	78	79	76	93	66	77.3	74
12.	Engine Room	86.3	84	84.3	82	82	78	81	79	92	66	77.4	74
13.	Air Grinder	84.3	80	84.8	77	77	84	82	73	96	64	75.3	72
14.	Typewriter	86.4	82	87.4	79	79	85	83	75	97	65	74.8	71
15.	TN+6	88.3	84	89.8	80	80	85	87	79	98	65	75.9	72
16.	Jet	93.8	89	94.3	82	82	94	90	81	106	70	79.0	76
	Range	26.5	15	18.5	10	10	25.0	19.0	14	20	14	10.5	10
	Mean	87.0	80.6	83.5	77.9	77.3	78.7	79.1	74.3	93.5	63.1	73.7	70.3
	Standard Dev.	7.4	3.7	4.7	3.0	3.0	5.2	4.8	3.7	5.2	3.2	2.8	2.8
	Rank	8	4	5	2	2	7	6	4	7	3	1	1
	Group	4	2	3	1	1	3	3	2	3	1	1	1

PURPOSE

It is the purpose of this paper to construct a Speech Interference (SI), frequency-weighting curve that can be used (1) to calculate a weighted SIL, (2) as a filter in a SLM, and (3) as a substitute for the NC type (NC, NCA, and ISO) contour at the 70 dB level. The curve will be designed to measure only the speech-interfering properties of noises. To the extent that speech interference is the determining factor in the judged loudness, annoyance, or office environment acceptability, this speech interference contour will measure that quantity.

Specifically, a contour will be developed that reduces the dispersion among the ratings of the 16 equally speech-interfering noises reported by Klumpp and Webster (2). The purpose will be to devise methods and means of better estimating the speech-interfering properties of noise without using the more involved AI technique.

SPEECH INTERFERENCE

On the basis of the results of Klumpp and Webster (2), the guidelines for developing a speech interference contour at the noise and speech levels used for those studies are clear. For 50 percent scores in relatively high level noises (as compared to acceptable offices), the frequency regions of the noise that limit the speech are centered at 500, 1000, and 2000 c/s. If the speech interference contour is to be used as a filter network in a sound level meter, sound of frequency below 300 c/s and above 3000 c/s must be discriminated against. Likewise, when used as a tangent-to-curve determiner the same frequency cutoffs must be observed. When used as a shaping network for calculating an SIL or average-curve-fitting within the octaves 500, 1000, and 2000 c/s, the center octave needs to be emphasized somewhat more than the others.

With these general guidelines a contour labeled SI-70 was developed as shown in figure 2. The shape of this contour is determined largely by the levels of the limiting spectra of the Klumpp and Webster (2) noises as shown in figure 5 of Webster and Klumpp (3). Using this SI-70 contour, the 16 equally speech-interfering noises of Klumpp and Webster (2) were rated as detailed in table 1. All of the ratings in table 1 (except those in italics which are taken directly from Klumpp and Webster (2)) are calculated measures, including those where it is assumed that the inverse of the SI-70 (and labeled A') is used as a filter network in a SLM.

In columns 1 and 2 of table 1 are listed the numbers of, and names of, the 16 noises. For comparison reasons the C and A weighting network ratings from Klumpp and Webster (2) are shown in columns 3 and 5, the NCA and ISO(R) ratings in columns 8 and 9, and the 3-band SIL in column 13. In column 4 is the rating that would result if a flat (C) weighting were used in a SLM but bandpassed to include only the octaves centered at 500, 1000, and 2000 c/s. This column is labeled C(R); the "R" specified here, as everywhere else in the table, "Restricted Range."

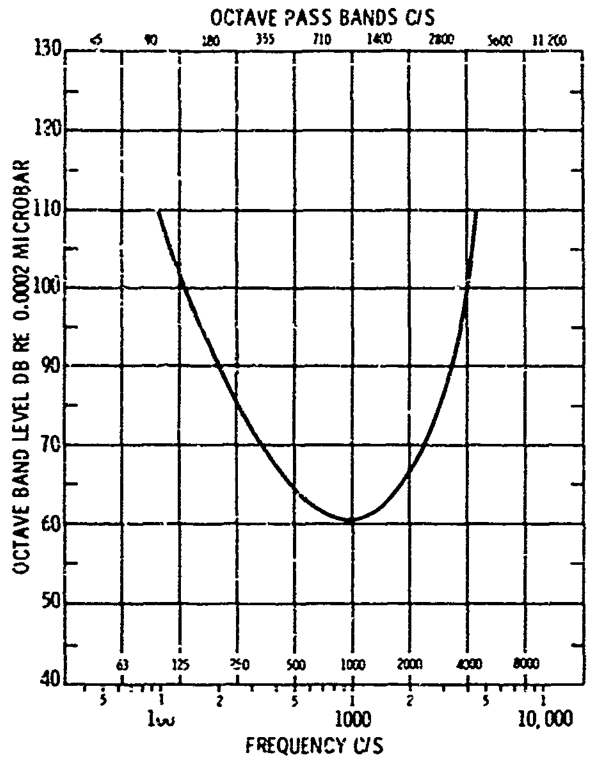


Figure VII-2. Speech Interference Contours.

The remaining columns in table 1 are ratings the 16 noises would get if the SI-70 curve were used as a new A network, namely A'; for the whole frequency range (column 6) or A'(R) for the restricted range. Column 10 lists results from using the SI-70 contours as the curve for the tangent-to-curve method. In column 12 are given the measures of the SI-70 curve when used as an averaging curve to find a five-octave SIL (based on center frequencies of 250, 500, 1000, 2000, and 4000 c/s). In column 14 are the results of restricting this averaging procedure to the usual restricted band (500, 1000, and 2000 c/s).

Below each column are two measures of dispersion: the range (highest minus lowest rating) and the standard deviation; and the mean rating on the 16 noises. The rank order refers to the relative smallness of the standard deviation. The smaller the standard deviation the better is that method in rating the noi :s to be, as they have been adjusted to be equally speech-interfering.

Some of the data in table 1 are plotted in figure 3 and reference to table 1 and figure 3 makes many points very evident. For example, the greatest reduction in the variation among the 16 noises occurs by merely restricting the bandwidth to the octaves centered at 500, 1000, and

2000 c/s (see columns 3 and 4 and the two uppermost curves). The new SI-70 is a considerable improvement over the A weighting when used as a filter in a SLM (column 5 vs. 6). Restricting the bandpass on the SI-70 as a filter doesn't improve things appreciably (column 6 vs. 7).

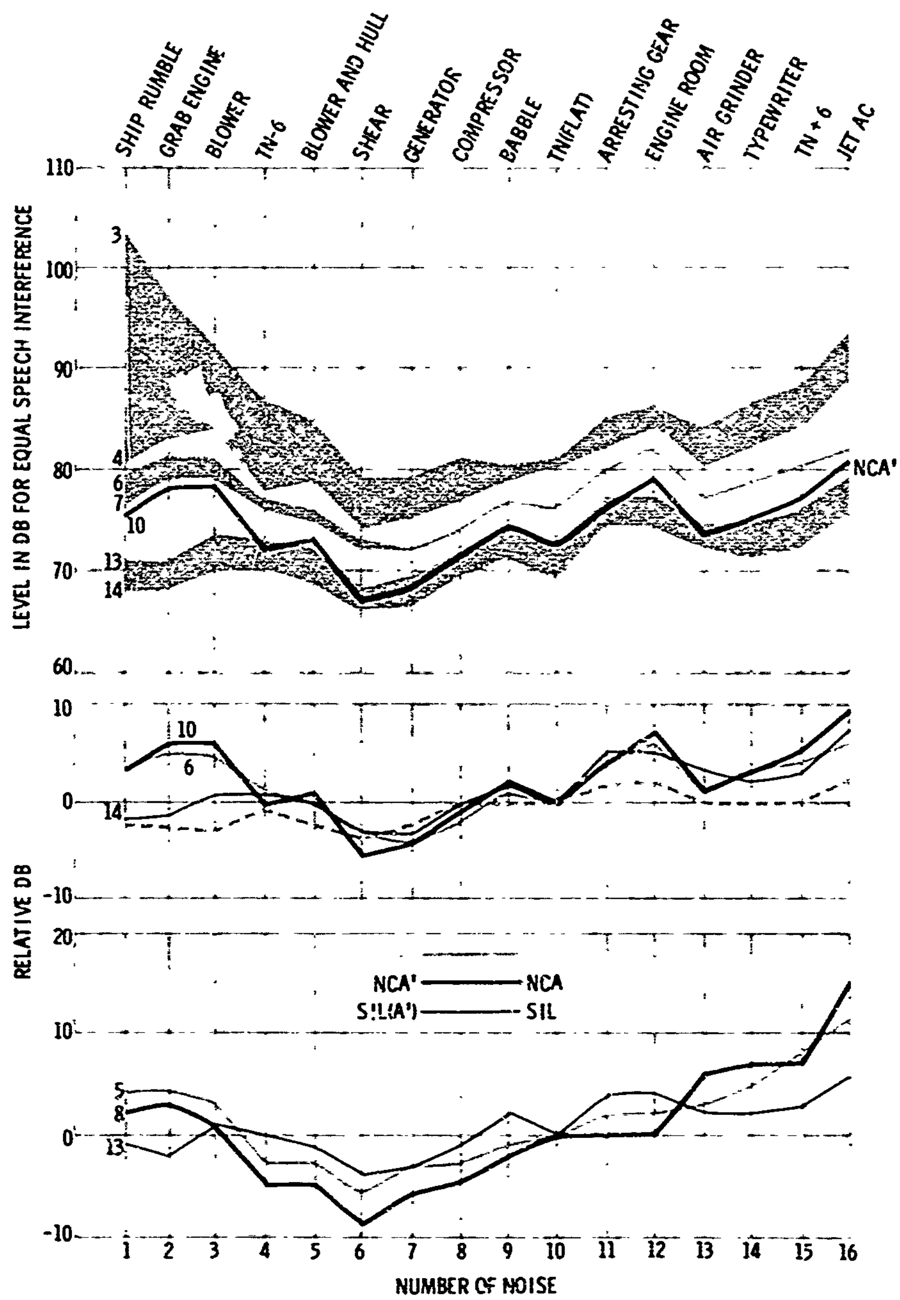


Figure VII-3. Ratings of 16 equally speech-interfering noises by various methods. Numbers at the left edge of the curves refer to column numbers in table 1.

When used as a Noise Rating curve the SI-70 contour (column 10) is markedly superior to the NCA (column 8), the ISO(R) (column 9), which is representative of both the NC and NCA curves when restricted in bandwidth, and to the loudness level calculation (column 11).

When used as an averaging curve to find a 5-band SIL, the SI-70 is good (column 12) but not as good as when finding a 3-band SIL (column 14) where it is as good as the conventional method (column 13). This latter finding is to be expected since, as pointed out earlier, in SIL calculations (as long as you stay with the same number of bands) the weighting should have no effect on any measure of dispersion. All the shaping does is add or subtract a constant number of dB from the original straight-line average.

In general, then, the SI-70 curve does what it was designed to do: It provides a single curve, which as far as predicting the speech-interfering properties of relatively high levels of noise, (1) makes a better filter network than the A weighting, (2) makes a better noise rating curve than the NC, NCA, and ISO curves, and (3) can be used as an averaging curve to find an SIL that is equivalent to the 3-band preferred frequency SIL.

A comparison among the best methods is detailed at the bottom of figure 3. In this plot selected columns are replotted from table 1 (or from the top of figure 3) but they are now equated on the thermal noise (TN), noise 10, and all ordinate values are relative. In the top set of four curves taken from columns 6, 10, and 14 of table 1, with the AI calculation from reference 3 added for comparative purposes, it is evident that the S[illegible] is the best simple predictor, and that the new SI curve w[illegible]ether used in a SLM (column 6, A') or as a noise rating curve (column 10, NCA') is slightly worse, especially for low frequency noises (noises 1, 2, and 3).

However, when compared to the old A weighting (column 5) or the old NCA (column 8), the new SI (or A') and NCA ' predictors are better, especiall[illegible] high frequency noises (noises 15 and 16).

On the basis of the theoretical improvements in measuring speech interference shown in table 1 and figure 3, an SI filter was constructed and added as an external filter to a Bruel and Kjaer model 2203 SLM. Figure 4 shows a comparison of the A weighting and the SI weighting on each of the 16 equally speech-interfering noises. As would be predicted from table 1 or figure 3, the difference between the two weightings is greatest on the high frequency noises (14, 15, and 16) and next greatest on the low frequency noises (1, 2, and 3).

Field trials on U. S. Navy ships are now in progress to determine if the new SI filter is sufficiently superior to the A weighting network in rating the capabilities for speech communication in noisy spaces to warrant its more general use.

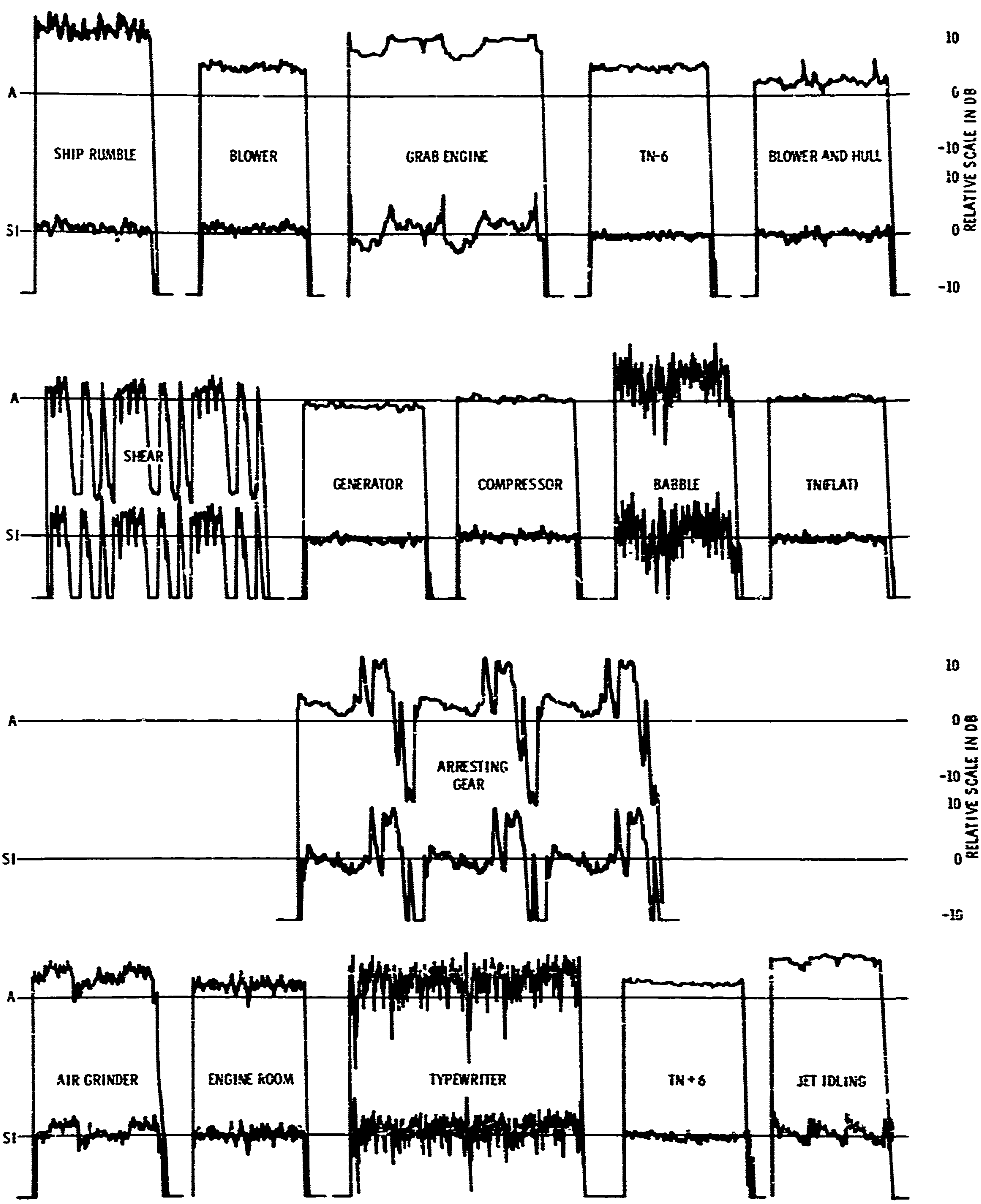


Figure VII-4. Tracings of 16 equally speech-interfering noises as seen via the A weighting filter network and the new SI weighting filter.

SUMMARY

This paper has attempted to delineate three different methods of rating noises. The three basic methods are: Sound Level Meter (SLM) readings using various frequency weighting networks; Noise Criteria Contours, where spectral peaks of noise become tangent to one of a family of rating curves; and average-level methods, the Articulation Index (AI) being the most sophisticated method and the Speech Interference Level (SIL) being the simplest to use.

Based on the data of Klumpp and Webster (2), a Speech Interference curve (SI-70, figure 2) was evolved. When this SI-70 curve was used as a frequency weighting network in a SLM, or as a noise-rating curve, or as a curve-fitting method of arriving at an SIL, it greatly reduced the spread of scores among the three measurement methods over the 16 equally speech-interfering noises of Klumpp and Webster (2).

ACKNOWLEDGEMENTS

I gratefully acknowledge the constructive criticism and continued interest of R. W. Young and B. S. Gales in this study, especially in the development of the mathematical equations for the rating methods. Likewise, without the earlier work of R. G. Klumpp and J. L. Leonard this work would not have been possible.

REFERENCES

1. Klumpp, R. G. and Webster, J. C., "Predicting Speech Interference From Physical and Psychophysical Measures of Ambient Noise," in Speech Communication Seminar, Stockholm, Sweden, 1962 Proceedings, September 1962

2. Klumpp, R. G. and Webster, J. C., "Physical Measurements of Equally Speech-Interfering Navy Noises," Acoustical Society of America. Journal, v.35, p.1328-1338, September 1963

3. Webster, J. C. and Klumpp, R. G., "Articulation Index and Average Curve-Fitting Methods of Predicting Speech Interference," Acoustical Society of America. Journal, v.35, p.1339-1344, September 1963

4. Kryter, K. D., "Methods for the Calculation and Use of the Articulation Index," Acoustical Society of America. Journal, v.34, p.1689-1697, November 1962

5. Beranek, L. L., Acoustics, p.419-420, McGraw-Hill, 1954

6. Peterson, A. and Bruel, P. V., "Instruments for Noise Measurements," p.16-13 and 16-14 in Harris, C. M., Handbook of Noise Control, McGraw-Hill, 1957

7. Beranek, L. L., "Revised Criteria for Noise in Buildings," Noise Control, v.3, p.19-27, January 1957

8. Netherlands. Research Center T.N.O. for Shipbuilding and Navigation Report 44S, Some Acoustical Properties of Ships With Respect to Noise Control, by J. H. Janssen, August 1962

9. Licklider, J. C. R. and Guttman, N., "Masking of Speech by Line-Spectrum Interference," Acoustical Society of America. Journal, v.29, p.287-296, February 1957

10. Stevens, S. S., "Procedure for Calculating Loudness: Mark VI," Acoustical Society of America. Journal, v.33, p.1577-1585, November 1961

11. Kryter, K. D., "The Meaning and Measurement of Perceived Noise Level," Noise Control, v.6, p.12-27, September-October 1960

12. National Aeronautics and Space Administration Technical Note D-1873, Some Effects of Spectral Content and Duration on Perceived Noise Level, by K. D. Kryter and K. S. Pearsons, April 1963

13. Young, R. W., "Don't Forget the Simple Sound-Level Meter," Noise Control, v.4, p.42-43, May 1958

14. Young, R. W., "Single-Number Criteria for Room Noise," Acoustical Society of America. Journal, v.36, p.289-295, February 1964

GENERALIZED SPEECH INTERFERENCE NOISE CONTOURS

J. C. WEBSTER

U. S. Navy Electronics Laboratory, San Diego, California

Based on an extensive literature review of the effects of noise on speech intelligibility, a series of noise rating curves are developed. These Speech Interference (SI) contours are intended to bridge the gap between (1) Noise Criteria (NC) and Alternate Noise Criteria (NCA) curves used to rate the suitability of offices, and (2) a Speech Interference (SI) noise rating curve that predicts the effects of higher level noises on speech intelligibility. The highest SI contour (SI-80) has a minimum at an octave level of 80 dB at 800 cps and is steeply sloped both above and below 800 cps. The lower level SI contours have minima at increasingly higher frequencies and have steep slopes for frequencies below the minima but gradually level off at the frequencies above the minima.

In another paper, Webster (1964) has developed a speech interference contour that best predicted the speech-interfering properties of the 16 equally speech-interfering noises described by Klumpp and Webster (1963). It is the purpose of this paper to generalize this single contour into a set of contours at higher and lower decibel levels. Ideally, these contours will extend the upper range of Beranek's (1957) Noise Criteria (NC) and Alternate Noise Criteria (NCA) curves for rating ". . . the maximum noise level at which office personnel feel they can accomplish their duties without loss of performance." Working spaces exist that exceed Beranek's (1957) maximum contour (NC or NCA-70) and that very often exceed his recommended maximum of NC-55. The rationale for developing these contours is that in certain spaces, certainly some shipboard areas, noise levels exceed NC-55 but work must and does continue, including voice communications. In these areas the major criterion must be acceptable speech intelligibility with little or no regard for loudness, annoyance, or comfort. To rate these spaces, therefore, contours based on comfort and speech communication performance must drop the comfort (loudness and annoyance) and base the rating and eventual acceptance only on those aspects of noise that affect speech intelligibility.

The contours to be developed then will ideally bridge the gap between (1) Beranek's (1957) NC and NCA curves that rate rooms on all aspects of noise, and (2) Webster's (1964) speech-interfering (SI) contour developed to predict the effects of noise upon speech intelligibility. This will be accomplished by utilizing the results of an extensive literature survey of the effects of noise and frequency bandwidth on speech intelligibility.

LITERATURE SURVEY

No discussion of the effects of various frequency regions of noise upon the masking of speech would be complete without considering the data of Miller

Reprinted from the *Journal of Speech and Hearing Research*
June 1964, Vol. 7, No. 2

NEL | A CONTRIBUTION OF THE NAVY ELECTRONICS LABORATORY, SAN DIEGO 52, CALIFORNIA

(1947). He used a broad-band (20-4 000 cps) noise and eight narrow bands of noise covering the range from 135 cps to 4 000 cps. He found that as compared to the masking done by the broad-band noise, (1) the low-frequency noise bands did virtually no masking at low levels and were very good maskers at high levels, (2) the high-frequency bands of noise were very effective maskers at low levels but didn't mask much more as the levels increased, and (3) at moderate to high levels, bands below 1 500 cps and especially those below 1 100 cps masked speech considerably better than those above 1 300 cps. His results can be summarized by saying that as the level of noise increases, or as the S-N (Speech-to-Noise) differential decreases, the masking effectiveness changes from higher frequency bands to lower frequency bands.

Miller's (1947) lowest band was from 135 to 400 cps. Dreher and Evans (1960-61) found that a masking band of noise from 50 to 300 cps never did decrease the intelligibility of speech, even at levels where a band from 600 to 4 800 cps masked speech phrases completely. They did find, however, that the 50-300 cps band, when added to the 600-4 800 cps band, caused an additional amount of deterioration in speech intelligibility.

Pickett and Kryter (1955), using sloped broad-band noises to mask speech, found some evidence to support Miller's (1947) findings based on narrow bands of masking noises. They found (see Figure 3 in Pickett and Kryter, 1955) that "low-frequency (LF) noise" (". . . —7 dB/octave at the low frequency end to —12 dB/octave at the high end . . . [as measured in octave-band, not spectrum, levels]") was not very effective in decreasing high levels of intelligibility but quite effective in further decreasing low levels. Similarly, they found (Figure 4, Pickett and Kryter, 1955) that 'high-frequency (HF) noise' (". . . slope of the HF . . . noise spectrum increased from +3 dB/octave to +5 dB/octave as frequency increased . . .") was quite effective in decreasing high levels of intelligibility but relatively less effective at decreasing lower levels.

Egan and Wiener (1946), doing the reverse of Miller (1947), found what bandwidths of speech were important to make speech intelligible in broad-band noises. Their data, which are summarized in Table 1, show that when speech is bandpass-filtered in broad-band noise, there is a frequency somewhere between 1 100 and 2 000 cps that is essentially the center or important frequency. As the width of the speech passband is increased more or less symmetrically around these center frequencies, intelligibility likewise increases. In general, the wider the band in octaves the better the intelligibility. For example, if 2 600 cps is subtracted from the top end of the 550-6 500 cps band, the intelligibility of the new 550-3 900 cps band drops a measurable amount (and the total bandwidth decreases from 3.56 octaves to 2.82 octaves). If the lower end of the band is extended downward by 210 cps (340 to 3 900 cps), the intelligibility and the bandwidth in octaves is again approximately equal to the original 550 to 6 500 cps band even though the bandwidth in cycles has been reduced from 5 950 to 3 560 cps.

TABLE 1. Intelligibility versus Bandwidth (from Egan & Wiener, 1946).

Word Intelligibility			Band Extent	Band Center	Bandwidth	
% at high S-N values	% at low S-N values	Rel. Rank	in cps	in cps	in cps	in octaves
88	63	1	130-9 200	1 093	9 070	6.14
81	46	2a	550-6 500	1 920	5 950	3.56
79	50	2b	340-3 900	1 150	3 560	3.52
77	42	3	550-3 900	1 460	3 350	2.82
70	30	4	550-2 500	1 170	1 950	2.18
65	25	5	870-3 900	1 840	3 030	2.16
53	18	6a	870-2 500	1 480	1 630	1.52
44	22	6b	550-1 500	908	950	1.46
42	12	7	1 300-3 100	2 007	1 800	1.25
22	10	8	870-1 500	1 120	630	0.79
20	5	9a	1 300-1 900	1 580	600	0.55
18	6	9b	1 800-2 500	2 120	700	0.47

It is interesting to note that in the three cases where the bandwidths in octaves were approximately equal (2a versus 2b, 4 versus 5, and 6a versus 6b) the bandwidth with the lower center frequency was more intelligible at low S-N values and in two of these cases (2 and 6) the reverse was true at high S-N values.

So the intelligibility of bandpass-filtered speech in broad-band noise increases as the width of the speech band increases in units of octaves around a center frequency of about 1 500 cps. This center frequency can be as low as 908 cps for bad conditions of speech in noise to a frequency as high as 1 920 cps for good conditions.

Some generalizations seem evident from the data of Miller (1947) and Egan and Wiener (1946). For speech to be very intelligible (80% nonsense syllable scores) in broad-band noise, the speech passband should be about 3.5 octaves wide and centered somewhere between 1 100 and 2 000 cps. Twenty per cent nonsense syllable scores remain for bandwidths between 0.5 and 0.75 octaves if centered between 1 100 and 2 000 cps. Low frequency bands of noise (below 1 100 cps and in bands of 1.6 octaves or less) do not mask broad-band speech well until they become relatively loud. High frequency bands of noise (above 900 cps and in bands of 0.75 octaves or less) mask broad-band speech somewhat at very low levels but do not mask speech appreciably more as the noise levels increase. At low intelligibility levels (less than 40% PB scores), narrow bands of frequencies above 900 cps (especially above 1 300 cps) do no additional masking whereas narrow bands of frequencies below 1 100 cps (especially below 700 cps, but, according to Dreher and Evans, 1960-61, not below 300 cps) do an appreciable amount of masking.

Can these trends be found in the data of others? The trends being that (1) a sufficient bandwidth for speech in noise is 3.5 octaves from 340 to 3 900 or

550-6 500, the center or important frequency being between 1 100 and 2 000 cps; (2) small amounts of high-frequency noise deteriorate speech intelligibility somewhat and greater amounts do no further damage; and (3) small amounts of low frequency noises won't deteriorate speech intelligibility at all but greater amounts can obliterate speech intelligibility.

The Articulation Index (AI), which is a measure that takes into account the difference in the spectrum level of speech and of noise over 20 contiguous frequency-limited bands, each of which contributes equally to speech intelligibility, yields a number that is a function of speech intelligibility. The AI as calculated by the Beranek (1947) method, see Kryter (1962a), confirms the bandwidth requirements for speech in noise; that is, the 340-3 900 cps band contains bands 2 through 18 of the 20 bands and the 550-6 500 cps band contains bands 4 through 20. Both bands include 17 of the possible 20 bands and would therefore yield an AI score of 0.85, which, according to Kryter (1962b), would permit a PB word score of 95%. Kryter (1962b, Figure 1) also shows that speech in the bands 0-600, and 1 200-2 400 cps is just as intelligible in the quiet as is speech in the bands 0-600, 1 200-2 400, and 4 800-9 600 cps. This indicates that speech frequencies above 4 800 cps do not contribute appreciably to intelligibility if there is sufficient speech energy in frequencies below 2 400 cps. In the same reference Kryter (1962b) shows that any bandwidth reduction within the 1 200-2 400 cps band reduces intelligibility appreciably, especially if the 0-600 cps band of speech is also eliminated. Kryter (1962b) found, as did Egan and Wiener (1946), that nonsense syllable intelligibility stays relatively high (60%) if only speech in the 1 200-2 400 cps band is passed. Unlike Miller (1947), Kryter (1962b) found that a band of high frequency noise (2 400-3 400 cps) does decrease intelligibility with increasing level even if the speech has already been low-passed at 1 700 cps. But like Miller (1947), he found that increasing levels of low-passed noise (200-1 100 cps) decreased intelligibility to lower levels than did the high-frequency band of noise even though the speech was already high-passed at 1 700 cps.

In another paper, Kryter (1960) presents more evidence on the important frequencies for speech when masked by noise. He was looking for three bands, each 500 cps wide, that would pass speech with maximal intelligence (and naturalness). He found that the lowest band should be centered at 500 cps (to 750), the next at 1 750 ±250 cps, and the third from 2 500 cps (to 3 000). In comparative tests using bands centered at 500, 1 500, and 2 500 cps he found the intelligibility to be better than when using contiguous bands of 1 500 cps-width, geometrically centered at either 400, or 1 000, or 1 580 cps. But he also found that among the 1 500 cps-wide bands the one from 100 to 1 600 (center freq. = 400) was appreciably better than the higher-centered one at 500-2 000 cps which was in turn better than the band from 1 000 to 2 500 cps. The superiority of the 100-1 600 cps band was especially evident at the least favorable (zero) speech-to-noise condition. This low band is four octaves wide (as compared to 2, and 1.32 octaves for the higher bands), but if the data of

Egan and Wiener (1946) are correct, the lowest 1.75 octaves, if not the lowest two octaves, could undoubtedly have been eliminated without any undue loss in intelligibility.

Pollack (1948) also has some data which show that the intelligibility of speech in noise is actually increased slightly if the speech frequencies below 350 cps are eliminated. He also confirms that speech frequencies above 3 950 cps add no more to intelligibility. Pollack (1948), who selectively high- and low-passed speech in noise backgrounds, found a ". . . shift in relative contributions to intelligibility from the low frequencies at low intensity levels [and low intelligibility scores] to the high speech frequencies at high levels. . . ." He found this shift ". . . by noting the frequencies at which one-half of the maximal contribution to the articulation index was made." These midpoints are, according to Pollack, ". . . 800, 1 010, 1 300, 1 430, and 1 620 cps . . . for levels of +10—, +20—, +30—, +40—, and +50—db orthotelephonic gain, respectively."

Dyer (1962), doing the reverse of Pollack (1948), namely, filtering the noise around a broad-band speech signal, found, like Pollack, that as the speech-to-noise differential increased, the frequency that divided the high- and low-passed noises into equal speech-interfering increments increased from about 1 000 to almost 2 000 cps.

If any generalizations can be made from the literature cited, they are: Speech frequencies below about 350 cps and above 3 900 cps are relatively unimportant to the intelligibility of speech in noise (Egan and Wiener, 1946; Kryter, 1960; and Pollack, 1948); according to Dreher and Evans (1960-61), noise frequencies below 300 cps are very ineffective in masking speech at tolerable listening levels unless higher noise frequencies are also present; according to Miller (1947), noise bands above 2 400 cps are very ineffective in masking speech; Kryter (1960) states that the most important narrow bands of speech energy are centered at 500 (to 750 cps), 1 750 ±250 cps, and 2 500 (to 3 000 cps); Kryter (1962b) also states that any decrease in speech bandwidth in the 1 200 to 2 400 cps band reduces intelligibility; the most important mid-frequency in broad-band speech, according to Egan and Wiener (1946), is somewhere between 1 100 and 2 000 cps and the bandwidth required for high intelligibility is about 3.5 octaves; Egan and Wiener (1946) also state that at good speech-to-noise conditions (at high levels of speech intelligibility) the important broad-band center frequency is around 2 000 cps and frequencies as high as 6 500 cps may be important; as the speech-to-noise conditions deteriorate, the important mid-frequency shifts down to around 1 000 cps and frequencies above 3 900 cps are ineffective (Egan and Wiener, 1946; Kryter, 1960; Pollack, 1948; and Dyer, 1962).

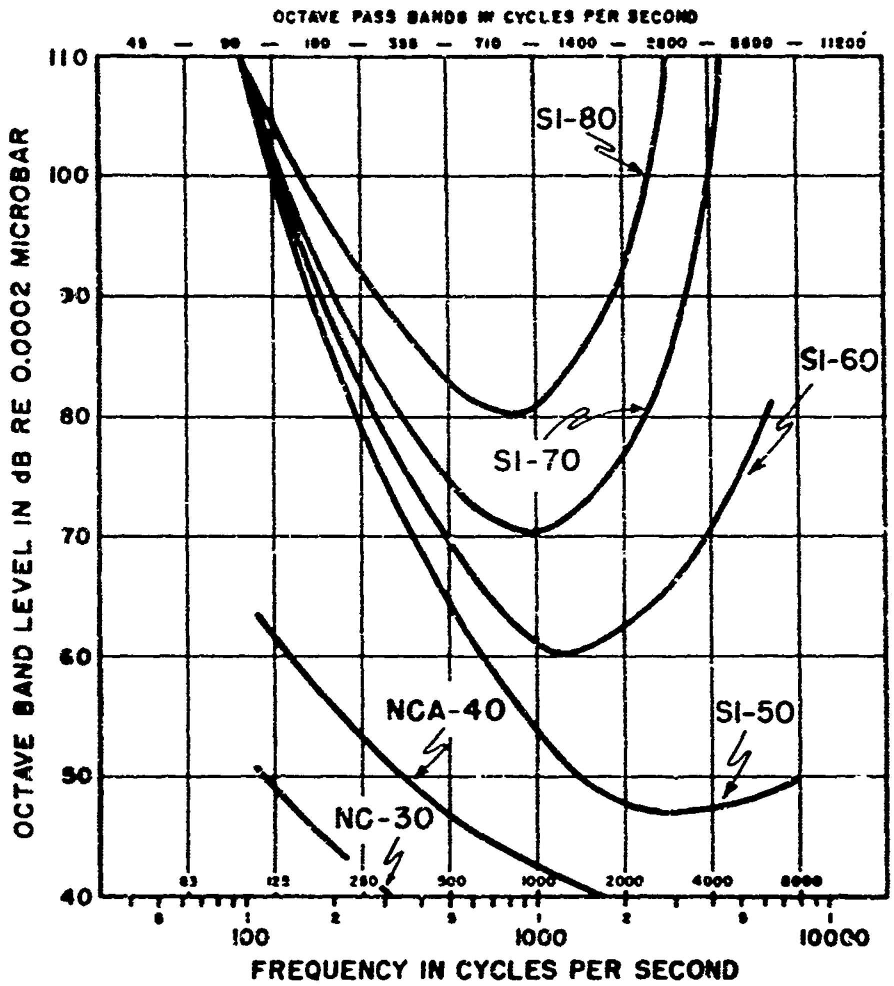


FIGURE 1. Speech interference noise rating contours.

SPEECH INTERFERENCE (SI) CONTOURS

The remaining SI contours (SI-50, 60, and 80) in Figure 1 were drawn relative to the SI-70 contour on the basis of the above observations, which are quantitative as regards frequency but only qualitative as regards sound pressure levels. The SI contours show (1) a gradual shifting from a minimum of 800 cps for SI-80 to 2000 cps for SI-50, (2) an increasing disregard of high frequency noise components from NCA-40 through SI-50, 60, and 70 to SI-80, (3) a sudden disregard of low frequency noise components from NCA-40 to SI-50, then increasing concern for low frequency noise for the contours SI-60, 70, and 80. These contours are developed on the basis that for levels of noise below the NC-30 contour, comfort, annoyance, and purely aesthetic values govern the use of a room. At NCA-40, Beranek (1957) states that all due allowance is made for the difference between loudness and speech interference, and above NCA-40 the environment is admittedly adverse and speech interference alone is hypothesized to be the determiner of acceptance.

In this regard it is interesting to note that in specifying the comfort of aircraft cabins (propeller-driven), Lippert and Miller (1951) define as "ideally quiet" a noise spectrum that becomes tangent to the SI-70 contour (between 500 and 1 000 cps). This spectrum is at least 15 dB above Beranek's (1957) NCA contour of 55 which he describes as "Very noisy; office environment unsatisfactory; . . . Not recommended for any type of office." Here is a case, and there are others, where the adaptability level of humans comes to their aid. A noise level that makes offices "unsatisfactory" is 15 dB less intense than a noise judged to be "ideally quiet" in airplane cabins. Lippert and Miller (1951) define a second contour exactly 20 dB higher as "quasi-comfortable." This letter level is 35 db above Beranek's "unsatisfactory office."

It is this adaptability feature of human behavior that gives rise to the rationale behind the discontinuity in the contours between NCA-40 (where comfort is of importance) and SI-50, and on through the SI-60, 70, and 80 contours, where the important aspect of the noise is its speech interference properties, not its loudness, its annoyance, nor its habitability and comfort properties.

SUMMARY

Because people do work, travel, and even go for entertainment (night clubs) in environments where noise levels greatly exceed "satisfactory office standards," it was thought desirable to extend Beranek's (1957) Noise Criteria (NC) curves to higher levels. In this extension, aspects of comfort, loudness, and annoyance are not as important as aspects of speech intelligibility. Therefore, a series of Speech Interference (SI) Contours have been developed based on what is known from the literature on the intelligibility of speech in noise.

These SI contours, which actually constitute the summary of this paper, show (1) an increasing disregard for high frequency noise components as the noise increases from levels of 40 dB to 80 dB (as estimated by an A-weighting network of a sound level meter), (2) a sudden disregard of low frequency noise components as the noise level passes 60 dB(A) and an increasing concern again for A levels above 80 dB, and (3) a shifting of the major concern for noise components centered at 2 000 cps for A levels of 40 dB and below to components centered at 1 000 cps and below in noises with A levels of 70 dB and above.

The opinions and assertions contained herein are the private ones of the writer, and are not to be construed as official, or as reflecting the views of the Navy Department or the naval service at large.

REFERENCES

BERANEK, L. L., The design of speech communication systems. *Proc. Inst. Radio Engrs.*, 35, 880-890 (1947).

BERANEK, L. L., Revised criteria for noise in buildings. *Noise Control*, 3 No. 1, 19-27 (January 1957).

DREHER, J. J., and EVANS, W. E., Speech interference level and aircraft acoustical environment. *Human Factors*, 2, 18-27 (1960-61).

DYER, W. R., The masking of speech by high- and low-pass noise. Tech. Document Rep. No. RADC-TDR-62-298, Rome Air Development Center, July 1962.

EGAN, J. P., and WIENER, F. M., On the intelligibility of bands of speech in noise. *J. acoust. Soc. Amer.*, 18, 435-441 (1946).

KLUMPP, R. G., and WEBSTER, J. C., Physical measurements of equally speech-interfering Navy noises. *J. acoust. Soc. Amer.*, 35, 1328-1338 (1963).

KRYTER, K. D., Speech bandwidth compression through spectrum selection. *J. acoust. Soc. Amer.*, 32, 547-556 (1960).

KRYTER, K. D., Methods for the calculation and use of the articulation index. *J. acoust. Soc. Amer.*, 34, 1689-1697 (1962a).

KRYTER, K. D., Validation of the articulation index. *J. acoust. Soc. Amer.*, 34, 1698-1702 (1962b).

LIPPERT, S., and MILLER, M. M., An acoustical comfort index for aircraft. *J. acoust. Soc. Amer.*, 23, 478 (1951).

MILLER, G. A., The masking of speech. *Psychol. Bull.*, 44, 105-129 (1947).

PICKETT, J. M., and KRYTER, K. D., Prediction of speech intelligibility in noise. Air Force Cambridge Research Center Tech. Rep. 55-4, June 1955.

POLLACK, I., Effects of high pass and low pass filtering on the intelligibility of speech in noise. *J. acoust. Soc. Amer.*, 20, 259-266 (1948).

WEBSTER, J. C., and KLUMPP, R. G., Articulation index and average curve-fitting methods of predicting speech interference. *J. acout. Soc. Amer.*, 35, 1339-1344 (1963).

WEBSTER, J. C., A speech interference noise rating contour. Submitted to *Acustica* for publication in 1964.

Received November 21, 1963

Reprinted from THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, No. 36, No. 9, 1662–1669, September 1964

Relations between Speech-Interference Contours and Idealized Articulation-Index Contours

J. C. WEBSTER

U. S. Navy Electronics Laboratory, San Diego, California
(Received 22 November 1963)

A comparison is made between speech-interference (SI) contours [developed from the speech-interfering properties of a representative sample of industrial (Navy) noises] and a set of articulation-index (AI) contours based on theoretical noise spectra encompassing the most-extreme spectra found among real noises. At a level of roughly 70 dB [as based on a speech-interference level (SIL) or decibel average over the octaves centered at 500, 1000, and 2000 cps], the SI and AI contours agree very well. At lesser levels of noise, the SI and AI contours purposely diverge because the SI contours were developed to bridge the gap between noise-criteria (NC) and alternate-noise-criteria (NCA) curves (developed by Beranek to rate both the annoying and speech-interfering aspects of office noises) and the basic SI70 (SI at 70 dB) curve. The complete set of AI contours points up, as do the SI contours, that as the ambient noise increases the importance or pivotal frequency shifts downward from around 2000 to around 1000 cps. The pivotal frequency is the frequency that divides the speech bandwith into two halves, each of which contributes equally to the total intelligibility. Some potential uses of the basic SI contour (SI70) as a filter network in a sound-level meter are discussed.

INTRODUCTION

IN an attempt to summarize the results of studies on equally speech-interfering noises,[1,2] a speech-interference (SI) contour was developed.[3]

This SI contour, a U-shaped contour centered at 70 dB at 1000 cps, shows the levels of noise in octave bands that best summarize the spectrum and levels of the original 16 equally speech-interfering noises.

In order to generalize this contour to lesser levels of noise, two contours centering on levels of 60 and 50 dB were interpolated[4] between this countour and the 40-dB alternate-noise-criteria (NCA) contour of Beranek.[5] And, to complete the generalization, a contour centered on 80 dB was extrapolated.[4] The generalizations represent an attempt to integrate the data on the 16 equally speech-interfering noises with other published data on the masking effects of noise on speech. These other studies are cited and interpreted in Ref. 4.

All of these SI contours, labeled SI80, 70, 60, and 50 are reproduced in Fig. 1. The SI contours at 80 and 70 dB represent the maximum octave-band levels of equally speech-interfering, quasi steady-state noises whose average level in the octaves centered at 500, 1000, and 2000 cps is approximately 80 and 70 dB, respectively. The SI contours at 60 and 50 dB represent a compromise between maximum octave-band noise levels for equally speech-interfering noises and for noises acceptable for office environments, where factors such as loudness and annoyance as well as speech intelligibility are important. The complete set of SI countours bridges the gap between speech-intelligibility predictors—especially the articulation index (AI)—and the noise- and alternate-noise-criteria (NC and NCA) curve that rate the acceptability of work spaces where speech intelligibility is but one of the important factors.

The rationale for the SI contour at the 70-dB level (and, subsequently, the generalized SI80-, 60-, and 50-dB contours) is detailed in Refs. 1–3, where it was pointed out that there are three ways in which noises have been rated in the past: (1) by measurement with frequency-weighting networks in sound-level meters, (2) by fitting the peaks of plotted noise spectra to

[1] R. G. Klumpp and J. C. Webster, "Physical Measurements of Equally Speech-Interfering Navy Noises," J. Acoust. Soc. Am. 35, 1328–1338 (1963).
[2] J. C. Webster and R. G. Klumpp, "Articulation Index and Average Curve-Fitting Methods of Predicting Speech Interference," J. Acoust. Soc. Am. 35, 1339–1344 (1963).
[3] J. C. Webster, "A Speech Interference Noise Rating Contour" (to be published).
[4] J. C. Webster, "Generalized Speech Interference Noise Contours," J. Speech & Hearing Res. 7, 133–140 (1964).
[5] L. L. Beranek, "Revised Criteria for Noise in Buildings," Noise Control 3, No. 1, 19–27 (1957).

families of noise-rating curves, and (3) by band-level averaging. It has been shown[3] that, if the SI70 contour is used as a frequency-weighting network (where it need be inverted), as a noise-rating curve, and as a weighting curve for calculating speech-interference levels (SIL's), the three methods give very similar results in predicting the speech-interfering properties of the 16 equally speech-interfering noises of Klumpp and Webster.[1]

This paper shows the relationships between the SI contours and a set of generalized AI contours, and discusses some of the merits, limitations, and uses of both the SI and AI contours.

I. DEVELOPMENT OF IDEALIZED ARTICULATION-INDEX CONTOURS

In a recent paper by Cavanaugh *et al.*[6] (CFHW) the essentials of the AI are displayed in a manner that makes calculations and manipulations relatively easy: "... the useful speech signal is shown as a dot field beginning at 200 cps and extending to 6000 cps. Each dot signifies a possible ½% contribution to the articulation index. The field is 30 dB 'high' and the greatest density of dots is at 2000 cps. The dot field is drawn for an average talker using 'conversational' speech effort" (Ref. 6, p. 481).

The CFHW plot, which they plotted in third-octave bands, has been redrawn on the basis of octave bands in Fig. 2. Also in Fig. 2 are drawn a series of contours, in 5-dB steps, of the idealized spectrum of a thermal

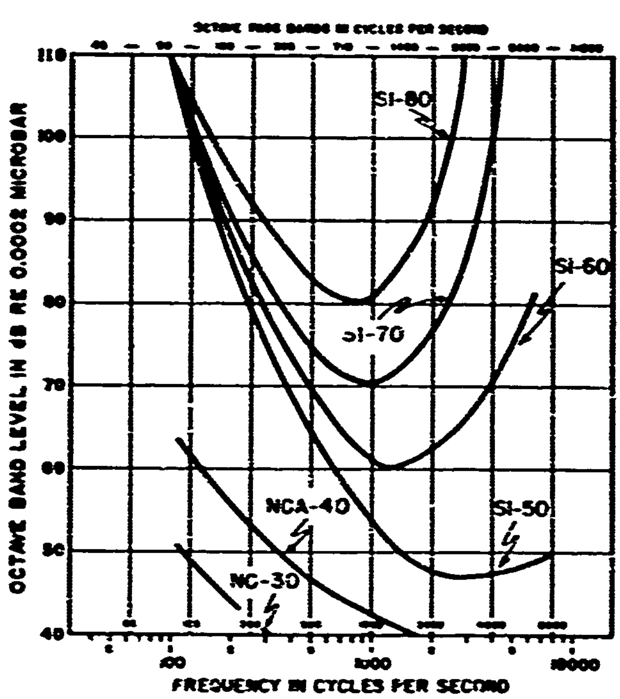


FIG. 1. SI contours (including the NCA40 and NC30 curves of Beranek[5]).

[6] W. J. Cavanaugh, W. R. Farrell, P. W. Hirtle, and B. G. Watters, "Speech Privacy in Buildings," J. Acoust. Soc. Am. 34, 475–492 (1962).

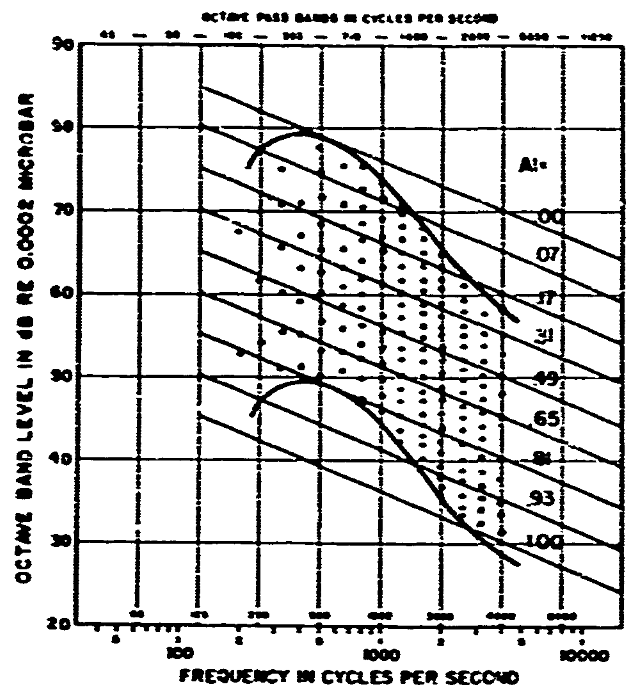


FIG. 2. AI speech region for "conversational-level" speech. The number of dots in each ½-octave band signifies the relative contribution of speech in that band to the AI. A series of idealized thermal noises with −6-dB/oct spectra are drawn in 5-dB steps. The number of dots above each noise contour is proportional to the AI of conversational-level speech in that level of noise. [After Cavanaugh *et al.*,[6] Fig. 5.]

noise with a minus 6-dB/oct spectrum slope. The number of dots between each two adjacent contours is proportional to the difference in AI. It will be noted that at an over-all level of 88 dB the AI equals zero. The 88 dB is derived by assuming that the noise band includes the octaves centered at 125, 250, 500, 1000, 2000, and 4000 cps and that the over-all level of a TN6 noise is 3 dB greater than the level in the octave with the greatest level—in this case, the 125 cps octave. AI's of 0.2, 0.5, 0.8, and 1.00 occur at over-all levels of 77, 68, 58, and 48 dB, respectively.

A similar relation between over-all level and AI can be determined by using the method shown in Fig. 2 with idealized noise-spectrum slopes of ±6, −9, and −12 dB/oct. From this type of information, Fig. 3 has been compiled. Figure 3 is a plot of the AI versus over-all noise level for various theoretical noises. Three different over-all bandwidths of noises are plotted on the abscissa: (1) to the extreme left is the over-all level of noise in the octave bands centered at 125, 250, ···, 4000 cps; (2) in the center is the over-all level of noise in the octave bands centered at 500, 1000, and 2000 cps; (3) and at the right is the over-all level of noise in the bandwidth from 600 to 4800 cps (or for octaves centered at 850, 1700, and 3400 cps).

It is evident from Fig. 3 that, within the general region of AI's from 0.2–0.8, all noises except the +6-dB/oct noise have the same general slope of AI score versus over-all levels of noise, regardless of the

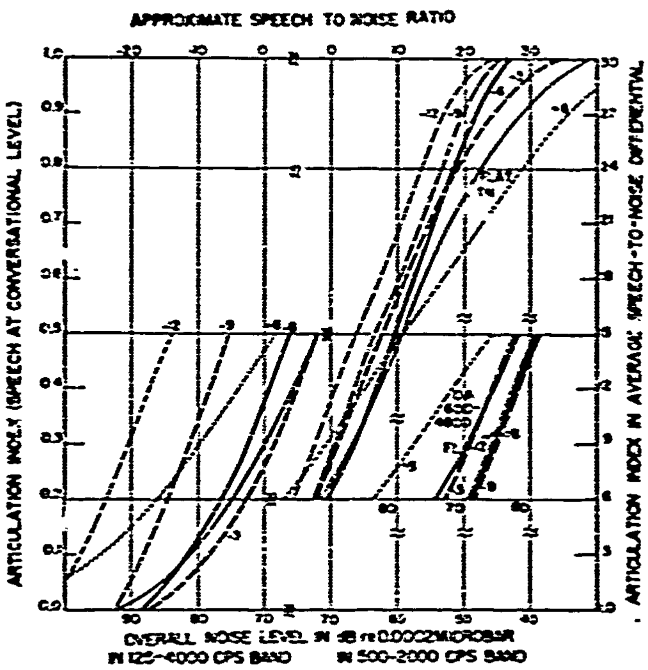


FIG. 3. AI as a function of the over-all level of noise for noises with idealized spectrum slopes of −12, −9, −6, flat, and +6 dB/oct (on spectrum-level basis) or ±9, −6, ±3 dB/oct (on octave-level basis). The over-all levels are based on bandwidths that include at the *left*: all the octaves centered from 125–4000 cps; *center*: octaves centered at 500, 1000, and 2000 cps; and to the *right*: octaves from 600–4800 cps (centered at 850, 1700, and 3400 cps).

bandwidth in which the over-all level of noise is measured. As compared to the other noise spectra, the +6-dB/oct noise shows a slower increase in AI with decreasing level of noise. Data to support this fact have been shown by Pickett and Kryter[7] in Fig. 1 of Ref. 2, where it can be noted that, for assumed speech levels of +10 and +20 above reference speech level, the AI's for the predominantly high-frequency type of noises (typewriter, TN +6, and jet) increase much less than for the other noises.

It is also immediately evident from Fig. 3 that reducing the bandwidth in which the noise is measured (from octaves centered at 125, 250, ···, 4000 cps to octaves centered at 500, 1000, and 2000 cps) reduces the spread in the noise levels for an equivalent AI value. It can also be seen that, if predominantly high-frequency-type noises (+6 dB/oct) are to be included in any predictive scheme, the bandwidth centered on octaves at 500, 1000, and 2000 cps is superior to the older SIL-calculation bandwidth of 600–4800 cps. Of course, since the majority of everyday noises predominate in low-frequency sounds (are not of the +6-dB/oct type), the 600- to 4800-cps band works well. The results shown in Fig. 3 suggest, however, that the general rule for predicting speech discrimination for speech masked by noise should emphasize the midfrequency regions (octaves centered around 500, 1000, and 2000 cps) and deemphasize frequencies outside of this region.

The CFHW dot-pattern plot is a method of displaying the perceptually important part of the statistical aspects of speech, as formulated by French and Steinberg[8] and by Beranek[9] from many listening tests. The importance of the 30-dB dynamic range, the 200- to 6000-cps bandwidth, and the greater contribution to speech intelligibility of speech frequencies around 2000 cps are amply displayed by the CFHW dot-pattern plot. The properties of noise that mask speech can be displayed by plotting in other forms data derived from plots like those in Figs. 2 and 3.

Figure 4, for example, is a plot of the spectra of the idealized noises at levels that yield an AI of from 0.00 to 0.05. That is, noises with spectra of −12, −9, −6, flat, and +6-dB/oct slopes at the levels shown in Fig. 4 will decrease the intelligibility of conversation-level speech to zero. If the over-all levels are lowered to the levels shown in Fig. 5, the AI is increased to 0.2, which, for conversational-level speech, would allow 50% of

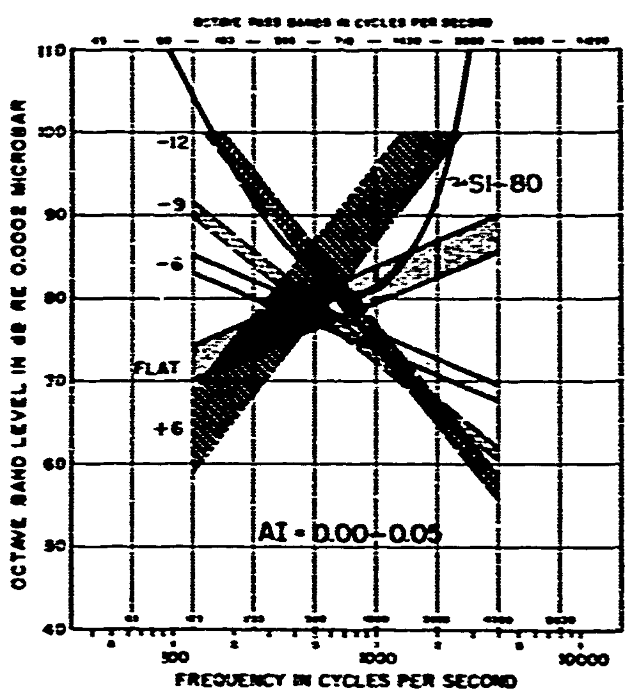


FIG. 4. Sound-pressure level in octave bands for an AI between 0.00 and 0.5 for conversational-level speech. The parameter is the slope of an idealized noise spectrum of −12, −9, flat, and +6 dB/oct (figured on a spectrum-level basis or −9, −6, −3, +3, and +9 dB/oct as plotted here on an octave-level basis). The over-all level of the noise can be estimated by adding 0.5 dB to the greatest octave-band level of the +9-dB/oct, 1.0 dB to the greatest octave-band level of the −6-dB/oct and 3.4 dB to the greatest octave-band level of the ±3-dB/oct (on octave-level basis) noises. Calculations are based on the Cavanaugh *et al.*[6] dot-pattern display of the AI. The SI80 contour from Fig. 1 has been superimposed for ease of comparison.

[7] J. M. Pickett and K. D. Kryter, "Prediction of Speech Intelligibility in Noise," AF Cambridge Res. Ctr. Tech. Rept. 55-4 (June 1955).

[8] N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," J. Acoust. Soc. Am. 19, 90–119 (1947).

[9] L. L. Beranek, "The Design of Speech Communication Systems," Proc. IRE 35, 880–890 (1947).

PB (phonetically balanced) words to be correctly recognized (according to Kryter,[29] Fig. 4). Similarly, successively decreasing the levels would result in an AI of 0.5 (Fig. 6), 0.8 (Fig. 7), and 0.95–1.00 (Fig. 8). Figure 9 is a composite of Figs. 4–8, with interpolation, to show the maximum level of noises of various monotonic spectrum shapes that would yield the designated AI's for speech at conversation level.

Because Fig. 9 is derived from theoretical and monotonic spectra noises does not limit its application to these kinds of noises. As has been pointed out earlier in this series of papers,[2] to get the best predictive value out of any set of rating contours, the contours should average through spectra that are irregular in shape. And, for predicting speech intelligibility, this average should center on the octaves centered at 500, 1000, and 2000 cps. A peaked noise spectrum, like voice babble or a diesel engine, should be fitted such that the point of the peak overshoots the contour enough so that the deviations of the noise spectrum from the noise-rating contour will average zero at the frequencies 500, 1000, and 2000 cps. When used in this manner, the AI contours in Fig. 9 can be used for noises that occur most commonly. And, of course, when used in this manner the contours are actually being used as a graphic means of finding a SIL based on octaves centered at 500, 1000, and 2000 cps.

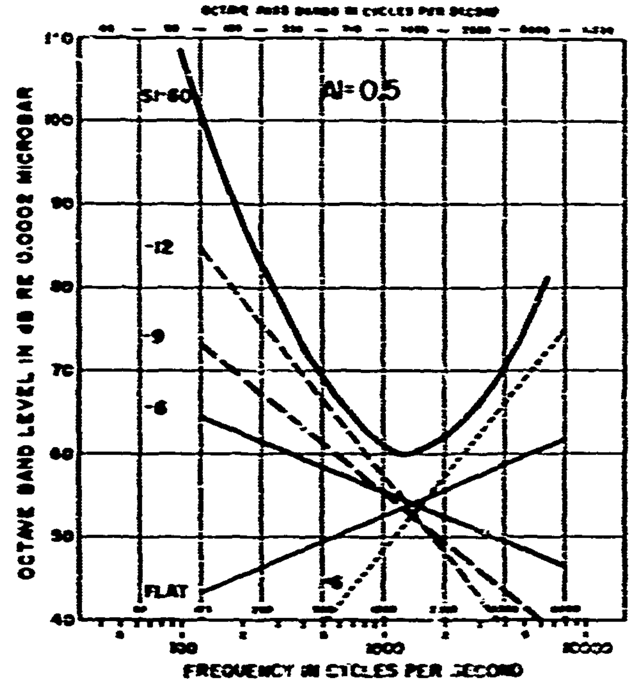


FIG. 6. Sound-pressure level in octave bands for an AI of 0.5 for conversational-level speech. The parameters are as detailed in Fig. 4. The SI60 contour from Fig. 1 has been superimposed for ease of comparison.

Regardless of the fact that they can be used on most noises, the shape of the contours in Figs. 4–9 are determined by the spectrum slope of the noises used to

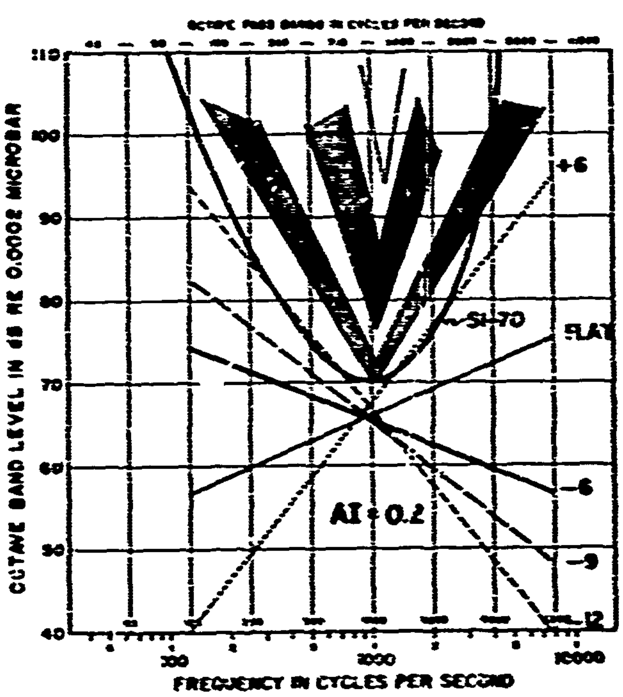


FIG. 5. Sound-pressure level in octave bands for an AI of 0.2 for conversational-level speech. The parameters are as detailed in Fig. 4. The SI70 contour from Fig. 1 has been superimposed for ease of comparison. In addition to the noise-spectrum slopes detailed in Fig. 4, slopes of ±20, ±30, ±40, and ±50 dB/oct (on octave-level basis) are added to this plot.

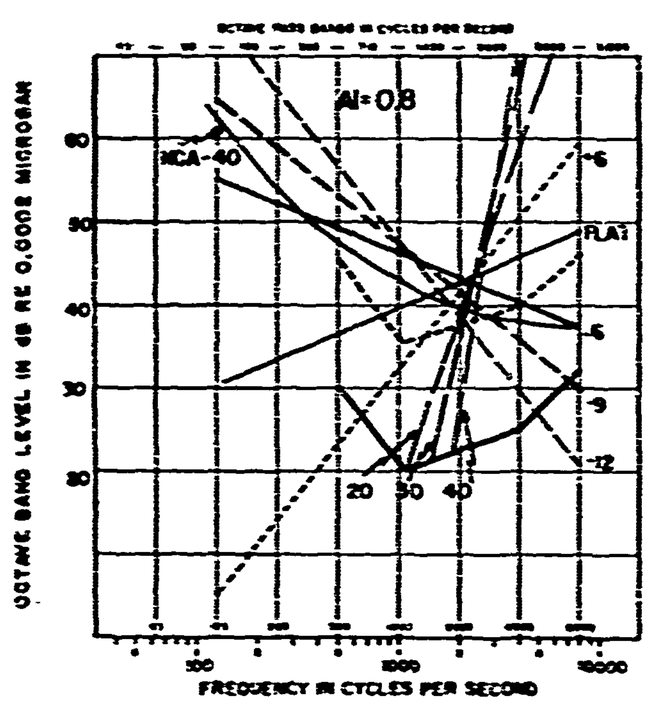


FIG. 7. Sound-pressure level in octave bands for an AI of 0.8 for conversational-level speech. The parameters are as detailed in Fig. 4. The NCA40 contour from Fig. 1 has been superimposed for ease of comparison. The bottommost curve is a conservative estimate of the SPL for the threshold of audibility.[21] The steeply sloping curves of +20, +30, and +40 dB/oct represent sharply filtered high-passed bands of noise that simulate high-frequency hearing losses.

[29] K. D. Kryter, "Validation of the Articulation Index," J. Acoust. Soc. Am. 34, 1698–1702 (1962).

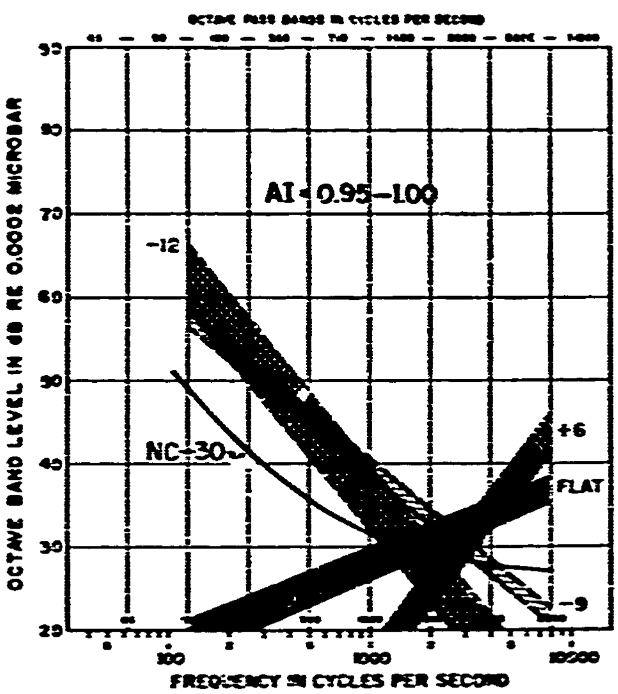


FIG. 8. Sound-pressure level in octave bands for an AI between 0.95 and 1.00 for conversational-level speech. The parameters are as detailed in Fig. 4. The NC30 contour from Fig. 1 has been superimposed for ease of comparison.

generate the data. Therefore, it should be considered whether the −12- to +6-dB/oct slopes encompass most commonly-occurring noises, and/or what happens to the contours if more- or less-extreme slopes are chosen. The −12- to +6-dB/oct slopes certainly seem to cover the majority of naturally occurring noises reported in the literature, at least those reported in the references in this paper. And, except for specific cases of sinusoids and narrow bands of noise, these spectrum slopes include most laboratory-generated noises.

However, to get an idea of what changes in contour shape would occur if more-extreme noise slopes were encountered, consider some of the details in Fig. 5 in which the gamut of slopes is covered. To define the limits: a high-frequency or high-passed noise with an infinite slope (vertical cutoff) would pass vertically through a frequency of 800 cps; an infinitely sloped low-passed (low-frequency) noise would pass through the frequency 2500 cps. That is, 20% of the dots in the CFHW dot-pattern plot lie at or below 800 cps or at or above 2500 cps. Slopes from roughly ±20 to ±40 dB/oct center between 1000 and 1250, as do the more gently sloped noises. So it is not so much the slope of the noise spectrum but the value of AI that determines the center frequency. Of course, ultimately it is the characteristics of speech that account for this, not so much the speech spectrum but the width (actually the narrowness) of the bands that contribute equally to the intelligibility of speech (the concentration of dots on Fig. 2).

On Fig. 9, it is apparent that as the AI increases from 0 to 1, the crossover (or minimum value) frequency shifts from about 630 to about 2500 cps. This again is not dependent on the limiting slopes of the extreme-noise spectra. Had the theoretical noises been limited to flat thermal noise and −6-dB/oct thermal noise, the same shift upward would have been evident. Note, for example, where the "flat" and "−6" noises cross on Figs. 5–7, representing AI's of 0.2, 0.5, and 0.8: namely, at about 1000, 1500, and 2000 cps, respectively.

Although noises with slopes of 30–50 dB/oct are not common, noise-induced hearing losses occur with such extreme slopes. And high-frequency hearing losses limit the hearing of low-level high-frequency speech sounds in much the same way as a high-passed masking noise. In Fig. 7, therefore, a series of 3 lines representing noise spectra that would produce simulated hearing losses at rates of from 30 to 50 dB/oct is drawn in at the position in frequency that would correspond to an AI of 0.8. A curve representing a conservative estimate of the sound-pressure levels for the threshold of audi-

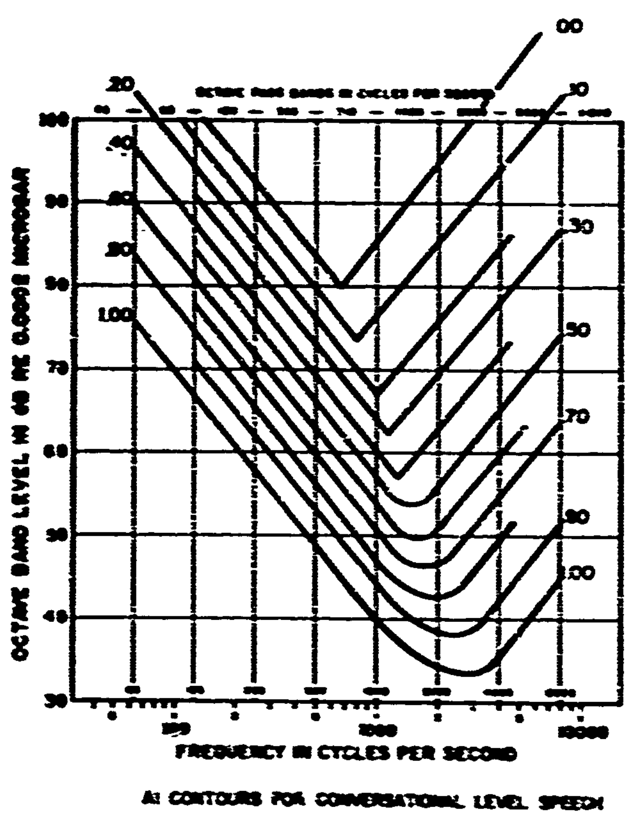


FIG. 9. AI contours for conversational-level speech, based on the dot-pattern AI display of Cavanaugh *et al.*,[4] and idealized noise-spectra slopes ranging from −12 to +6 dB/oct (on a spectrum level basis) of a ±9-dB/oct band, based on octave-level basis. These contours will estimate very accurately the AI for monotonically increasing or decreasing noise slopes, but will be in error for noises having either peaks or valleys in their spectra at or near the valleys in these contours. These errors can be minimized considerably, however, if the peaks and valleys of the noise spectrum are adjusted to "average through" the appropriate contour; see text for more detail. For detailed spectra of noises, these contours will fit; see Figs. 4–8.

bility[21] is also drawn in Fig. 7 so that the 30- to 50-dB/oct slope can be roughly interpreted in terms of hearing level (i.e., loss from normal hearing). To the extent that hearing level can be simulated by masking noise of comparable level, it can be seen that the AI scores of 0.8 can theoretically be achieved by persons with (1) normal hearing up to 1000 cps and simulated hearing losses of 15 dB at 2000 cps and 28 dB at 4000 cps, or (2) normal up to 1500 cps and losses of 10 dB at 2000 cps and 45 dB at 4000 cps, or (3) (by definition) normal up to 2500 cps and an infinite loss above 2500 cps.

Also drawn in Figs. 4–8 are the SI or the NCA or NC curves from Fig. 1 that most nearly approximate the given level of AI. Note, for example, that the SI80 and 60 correspond roughly to AI's of 0.0 and 0.5 and that SI70 corresponds quite closely to an AI of 0.2. The agreement of NCA40 to an AI of 0.8 and NC30 to an AI1.00 is not nearly as good and it was not intended that they should agree. These latter NC contours are for rating the acceptability of offices where speech intelligibility is only one aspect of the total noise environment.

In summary, it is quite evident that the SI contours (derived from experimental results) and the AI contours (derived from using idealized noise spectra and the CFHW display of the AI formulation) agree in many details; the agreement would be improved by interpolating a few more SI contours. This should not be too surprising, inasmuch as the AI formulation as well as the SI contours were based on experimental listening results. The degree of agreement could be improved by interpolating a few more contours in the region between the NC30 and the SI60 contours, by reexploring the statistics of occurrence of naturally occurring noises, or by essentially rerunning the 16 original Klumpp and Webster[1] noises at other levels of speech and/or noise and/or intelligibility. The agreement is not intended to be perfect in any case because the SI50 and 60 contours are meant to be interpolations between the NC30/NCA40 curves [for an ideal (NC30) or nearly ideal (NCA40) environment] and the SI70 contour known to predict speech intelligibility well at relatively high (for offices) noise levels. However, the agreement, as far as it now goes, between the AI and SI contours shows that the AI formulation (based solely on the filtering of speech) does hold within certain limitations when applied to noise-masked speech.

II. DISCUSSION OF THE ARTICULATION INDEX

The next question is, "What are the strengths and weaknesses of the AI?" Kryter[19] has shown that, within the data of any one experimenter (or any one group of experimenters), the AI very nicely takes

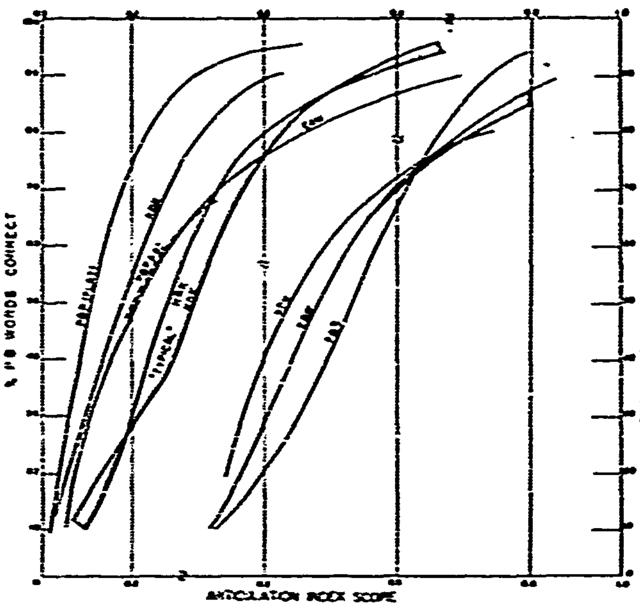


FIG. 10. Nonsense syllables and PB word scores for given AI's from the data of J. M. Pickett and I. Pollack [J. Acoust. Soc. Am. 30, 955–963 (1958)], Kryter,[19] Miller [Psychol. Bull. 2–44, 105–129 (1947)], J. P. Egan and F. M. Wiener [J. Acoust. Soc. Am. 18, 435–441 (1946)], and French and Steinberg,[4] as summarized by Kryter.[19]

account of the physical levels of filtered but otherwise unprocessed speech and noise and reduces to one parameter (AI) a monotonic, low-variance, band-limited predictor of word (or syllable, sentence, etc.) intelligibility. However, he also collects into one paper[20] the data to show that for any given AI the PB word score or nonsense-syllable score varies over a 40% range (see Fig. 10).

Earlier, Licklider[22] had also summarized the discrepancies between experimental data and AI prediction; he stated, "When comparisons were made between two speech communication systems with equal (computed) articulation indices, one having high speech-noise ratio and low bandwidth and the other low speech-noise ratio and high bandwidth, the wide-band system usually turned in the better measured performance." Concerning the important, or center, frequency for high- and low-pass filtered speech in the quiet (which is variously listed as being between 1600 and 1900 cps and is the frequency that divides the AI into two equal parts), Licklider reiterates (from the data of Pollack[23]) what has already been demonstrated clearly in this paper and in the Klumpp and Webster paper,[1] on which it is based, that, "In the tests in which masking and filtering were combined, the high-pass and low-pass functions did not show the kind of symmetry just described: namely, around a certain

[21] J. C. R. Licklider, in *Handbook of Experimental Psychology*, S. S. Stevens, Ed. (John Wiley & Sons, Inc., New York, 1951), Chap. 25, p. 995, Fig. 5.

[22] J. C. R. Licklider, "Three Auditory Theories," in *Psychology: A Study of a Science*, S. Koch, Ed. (McGraw-Hill Book Co., Inc., New York, 1959).

[23] I. Pollack, "Effects of High Pass and Low Pass Filtering on the Intelligibility of Speech in Noise," J. Acoust. Soc. Am. 20, 259–266 (1948).

frequency of 1600 to 1900 cps. The discrepancy was marked." The point of symmetry shifted successively downward in frequency as the speech-noise ratio deteriorated. Concerning the relation between AI and different speech materials, Licklider states, "We have been assuming that the articulation score for each category of speech material had its own fixed relation to articulation index.... But when Hirsh, Reynolds, and Joseph[14] plotted their word scores against their nonsense syllable scores, they found that they had to draw separate curves not only for the various categories (numbers of syllables) but also, within each category, for filtering and masking." Finally, concerning the relation between AI and different spectrum noises, Licklider quoted the results of Pickett and Kryter[7] to the effect that for four different noises the intelligibility scores for equal AI scores fell along four different curves.

Briefly resummarized, the deficiences of the AI are that: the frequency and level bands are not linearly additive [as Licklider[12] states, "...short fat (low S/N, wide bandwidth) regions...yield higher scores than tall thin (high S/N, narrow bandwidth) regions of equal area..."]; equal AI scores yield different speech scores dependent upon (1) the types of speech materials[14] and (2) the spectrum of the noise[1,7]; and the "importance frequency" lowers with adverse listening conditions.[13]

Although Kryter[15] has added some corrections to minimize certain of the deficiencies of (and to extend greatly the usefulness of) the AI, certain deficiences are still inherent. It is, therefore, not a condemnation of any proposed system of estimating speech intelligibility that it does not agree completely with the AI calculation. In any case, in the region of 0.2 AI where the SI70 (SI contour at the 70-dB level) was based on direct data,[1,3] the agreement between AI and SI is very good (see Fig. 5). Of course, part of the agreement is because the theoretical noises chosen to plot the AI contours (Figs. 2, 4–9) encompass the spectra of the real noises of Klumpp and Webster[1] upon which the basic SI contour at the 70-dB level (SI70) was developed. And the relatively large discrepancies between the AI and the SI and/or NC or NCA curves at lower levels of noise in Figs. 6–8 represent a deliberate compromise between the speech-interfering aspects of the noise and its loudness and annoyance aspects!

III. POTENTIAL USES FOR THE SPEECH-INTERFERENCE CONTOURS

It is too early to determine the usefulness and limitations of the SI contours. A few uses have been found. For example, a frequency-limiting network has been built to the inverse specifications of the SI70 contour. When used in a sound-level meter to measure the 16 equally speech-interfering noises of Klumpp and Webster,[1] it gave more-homogeneous measurements than any existing sound-level-meter filter network.[3] Or, restated, it appears to be a good network to insert into a sound-level meter if the SI effects of noises are to be measured.

A pencil and paper calculation has been made to determine if this "speech interference" filter used in conjunction with the already existing "A" network in sound-level meters might predict the acceptability of the noise environment in a potential office. This calculation used the noises and rationale of Beranek[5] in the development of the NC contours. In his paper, Beranek (Ref. 5, p. 23) concludes that, "A majority of the [office] personnel object to a noise whose octave-band levels at the low frequencies exceed, for a given SIL [speech-interference level, decibel average of octaves in the 600- to 4800-cps range], the values given by the NCA curves, even when the SIL is low...objections occur whenever the LL-minus-SIL difference exceeds 30 units."

These observations are supported by data in his Fig. 5 (Ref. 5), which are summarized in this paper in Table I, columns 1–5. Column 1 lists Beranek's 8 noises. In column 2, a **B** represents a noisy room "before" sound treatment; A designates "after." Columns 3–5 are taken directly from Beranek's Fig. 5 and superscript "a's" indicate rooms judged unsatisfactory before (LL–SIL$\geq$30) but satisfactory after (LL–SIL$\leq$22), on the basis of the LL-minus-SIL criteria. In columns 6 and 7, calculations based on the spectra given by Beranek[5] are made for A weighting (column 6) and SI70 weighting (column 7). That is, columns 6 and 7 are the best estimates of what a sound-level meter using A and SI70 weightings would have measured. Note that in general if the difference between columns 3 and 4—namely, column 5– is equal to or greater than 30, the difference between columns 6 and 7—namely, column 8—is equal to or greater than 5. It would appear, therefore, that if a sound-level meter had two weighting networks available, A and SI70, a preliminary judgment could be made as to the acceptability of the noise environment of a room by noting the difference in reading between the two.

Beranek (Ref. 5, p. 24) warns, "*It is not recommended that A-scale readings be used in specifications because the same A-scale reading may be obtained for a wide variety of shapes of spectra. Furthermore, the eight octave bands are necessary in the engineering design of noise control measures and no single number can substitute.*" By using the difference between an A and an SI70 scale, some of Beranek's objections to a single number are overcome. And, since the SI70 scale discriminates against both low and high frequencies more than the A scale, some of Beranek's objections *re* "... the same...reading...for a wide variety...of spectra..."

[14] I. J. Hirsh, E. G. Reynolds, and M. Joseph, "Intelligibility of Different Speech Materials," J. Acoust. Soc. Am. 26, 530–539 (1954).

[15] K. D. Kryter, "Methods for the Calculation and Use of the Articulation Index," J. Acoust. Soc. Am. 34, 1689–1697 (1962).

can be discounted. Both high- and low-frequency sounds are judged to be "annoying," though they are not particularly "speech interfering." There can be no objection to Beranek's statement as to the necessity of a complete eight-octave spectrum for noise-control purposes; but, for preliminary survey purposes, potentially objectionable noise environments may be found by sound-level-meter measurements alone if an SI70 scale is added (perhaps to replace the B scale).

Beranek (Ref. 5, p. 24) also points out the obvious fact that "Complaints were also registered (room f) [Ref. 5, Table I, noise f] when the LL-minus-SIL difference was less than 30 units but the SIL [600–4800 cps) exceeded acceptable values." Noises b and f marked by superscript b are examples of this, and acceptable improvements occurred when the SIL values (column 4) were substantially reduced. The same improvement could be measured by substantial reductions in A of SI70 scale readings (columns 6 and 7) or by an SIL based on the octaves 300–2400 cps (column 9).

It is interesting to note in comparing columns 4 and 9 (SIL's based on 600–4800 vs 300–2400 cps) that on the average the 300- to 2400-cps SIL's are 5.25 dB greater than the 600- to 4800-cps SIL's, which implies that "typical" office noises have a spectrum slope (when measured in octave bands) of about −5 dB/oct in the speech range from 300–4800 cps. And, significantly, the acceptable ones slope −4.5 dB/oct and the objectionable ones, −6 dB/oct. This 5-dB difference between the 300- to 2400- and 600- to 4800-cps SIL is not typical of all noises. The difference was only 1.5 dB for the 16 ship noises (Ref. 1, Table 2) but was 2.5 dB when a weighting was added to correct for how often the 16 noises probably occurred in a ships' environment.

Other comparisons on Table I show the SI70 weighting to read on the average 1.6 dB higher than the SIL (600–4800 cps) and 3.6 dB lower than the SIL (300–2400 cps), which makes it a pretty good prediction of the recommended compromise SIL based on octaves centered at 500, 1000, and 2000 cps.[1,2]

Using the data of Table I as a first approximation for setting criteria using the A and the SI70 network, it would appear that the noise environment in an office will be judged unsatisfactory if the sound-level-meter reading using the SI70 filter network is 50 dB or above or if the difference between the A reading and the SI70 reading is 5 dB or above.

TABLE I. Noise ratings in dB for Beranek's [Noise Control 3, No. 1, 19–27 (1957)] eight office noises. LL: loudness level. SIL: speech-interference level. A, SI70: calculated sound-level-meter readings when using an A or an inverse SI70 frequency-weighting network.

1	2	3	4	5	6	7	8	9
NOISE		LL	SIL6-48	Δ	A	SI70	Δ	SIL3-24
a	B	73	42	31[a]	49	44	5[a]	47
a	A	64	41	23	46	43	3	45
Δ		9	1		3	1		2
b	B	71	48[b]	23	52	50	2	51
b	A	62	40	22	45	41	4	45
Δ		9	8		7	9		6
c	B	64	29	35[a]	41	31	10[a]	38
c	A	46	25	21	30	26	4	30
Δ		18	4		11	5		8
d	B	59	29	30[a]	37	32	5[a]	34
d	A	50	28	22[a]	35	31	4	31
Δ		9	1		2	1		3
e	B	69	34	35[a]	40	34	6[a]	38
e	A	55	31	24	36	32	4	36
Δ		14	3		4	2		2
f	B	82	55[b]	27	60	55	5[a]	60
f	A	73	47	26	52	48	4	52
Δ		9	8		8	7		8
g	B	75	41	34[a]	55	44	11[a]	51
g	A	57	32	25	37	33	4	35
Δ		18	9		18	11		16
h	B	57	35	22	40	37	3	41
h	A	54	30	24	37	31	6	36
Δ		3	5		3	6		5

[a] Judged unsatisfactory before and satisfactory after treatment.
[b] Judged unsatisfactory (because SIL is too large).

Until such times as some data are collected, using the whole series of either the AI contours or the SI contours for rating noises, no statement can be made as to their practical usefulness. The value of their use in better understanding the relationships between noise spectra and levels and the masking of speech has been the purpose of this presentation.

ACKNOWLEDGMENTS

Grateful acknowledgment is made to R. S. Gales and R. W. Young, who have offered constructive criticism in all phases of this study, and to R. G. Klumpp and W. E. Green, whose painstaking work on the original experiment makes any conclusions on these further analyses more meaningful.

Important Frequencies In Noise-Masked Speech

J. C. WEBSTER, PhD
SAN DIEGO, CALIF

Introduction

When is a hearing loss an impairment, a handicap, or a disability? What is the difference between impairment, handicap, and disability? When is monetary compensation due if the hearing loss has been caused by military or industrial noise? These problems were discussed at the October, 1963, meeting of a symposium sponsored by the National Research Council-Armed Forces Committee on Hearing and Bio Acoustics (CHABA), which was chaired by Hallowell Davis.

Many things were discussed, but Dr. Davis tried to bring some order out of chaos by proposing that hearing "impairment" be considered the least noxious descriptor and be used to define a "defective function." He then defined a hearing "handicap" as being sufficient to "reduce one's efficiency in daily living" or put one at a social disadvantage. And he defined a hearing "disability" as the most noxious and sufficient "to reduce one's earning power."

Whether compensation was payable upon the discovery of an "impairment," proof of a "handicap," or only when a job change was necessary ("disability") was not discussed directly since these were medical men and audiologists, not lawyers and industrial (labor-management) representatives.

The question of how best to measure and quantify the degree of hearing impairment for speech was discussed loud and long with two points of view being represented: the present method, and a proposed change sponsored by Dr. Karl Kryter.

Dr. Davis traced the history of the present method of calculating "hearing impairment for speech" from the average of pure tone losses at 500, 1,000, and 2,000 cps. The total history behind this system dating back to Dr. Harvey Fletcher and Dr. Edmund Fowler will not be belabored here. Suffice it to say that the present method was ar-

Submitted for publication Dec 24, 1963.
United States Navy Electronics Laboratory.

rived at by the Committee on Conservation of Hearing sponsored by the American Academy of Ophthalmology and Otolaryngology (hereafter referred to as AAOO). The AAOO committee started on the premise that "Because of present limitations in speech audiometry, the hearing level for speech should be estimated from measurements made with a pure tone audiometer." [1]

Dr. Kryter presented the data from two of his recent papers [2,3] to support his reasons for a change in calculation method. In his recent paper,[2] Kryter recommended "that the impairment due to noise-induced hearing loss for the understanding of speech be estimated from the average hearing level at 1,000, 2,000, and 3,000 cps." Remember that the present (AAOO) method uses the average pure-tone hearing level at 500, 1,000, and 2,000 cps as the basis for determining impairment for everyday speech under everyday listening conditions,[1] for all types of hearing losses.

Kryter's Rationale

The actual content of Kryter's CHABA presentation will not be repeated here; it is well documented in two papers.[2,3] Only his hypotheses and rationale will be discussed.

Kryter's [2] hypothesis is that evidence for the 500-2,000 cps "fence" is based on *speech threshold tests* (usually called Speech Reception Tests and abbreviated SRT), whereas it is *speech discrimination tests* (usually measured with phonetically balanced word lists and abbreviated PB) which are important. Actually it is neither PB discrimination nor SRT that is the crux of the matter, but, as the AAOO committee states in its report, "[it is] the ability to hear sentences and repeat them in a quiet environment [that] is taken as satisfactory evidence of correct hearing for everyday speech." There is no argument that speech (PB) discrimination at supra-threshold levels, and especially in a background of noise, is a far more valid test of speech intelligibility than is speech threshold, but neither measures "everyday speech" in "everyday conditions."

Kryter [2] stated that his ". . . recommendation [of a 1,000, 2,000, and 3,000 cps fence] has the support of related studies [of Mullins and Bangs [4] and French and Steinberg [5]] and is contested only by those experiments [by Quiggle, Glorig, and Summerfield,[6] and Harris, Haines, and Meyers[7]] in which . . . threshold . . . tests . . . for . . . words . . . were used."

The discussion at the CHABA symposium took four forms: criticism of Kryter's paper, presentation of contrary evidence, a discussion of where on the impairment-handicap-disability scale the results of all studies fell, and how indeed does one measure "handicap" in particular.

This paper is not intended to be a complete transcript of the CHABA symposium so only two aspects of the discussion will be pursued: Some general criticisms of Kryter's experiment and some data of my own (which were also presented at the CHABA symposium).

Speech Discrimination vs Speech Threshold

One of Kryter's major hypotheses is that PB tests are more typical of, or at least better predictors of "everyday speech in everyday environment" than are SRT tests. In a factorial analysis study of speech perception, Hanley [8] did indeed find separate factors for thresholds (including 500, 1,000, 2,000, and 4,000 cps tones and sentences, spondee, nonsense syllable and PB speech) and for "resistance to distortion" and for "resistance to masking." Both of Hanley's "resistance to-" factors included tests which made "normal ears" hear speech as it might be heard by partially deafened individuals. Solomon, Webster, and Curtis [9] also found separate speech threshold factors, a "distortion or masked" factor, and a distraction (selective attention) factor. Peculiarly enough, the speech threshold factor included sentences, spondees, and nonsense syllables but not PB words, whereas the masking factor included PB words, but not spondees (the basis of SRT tests). Neither the Hanley[8] nor the Solomon, Webster, and Curtis[9] studies had hard-of-hearing subjects, but the test batteries included tests which temporarily made normal hearing subjects hear as if they were partially deafened in one form or another.

In studies where both normal and hard-of-hearing subjects were used, somewhat different results obtain. Ross, Huntington, Newby, and Dixon[10] studied the relationships between pure-tone audiometric measures and PB scores for groups of normal and sensorineural subjects. They included tests of difference limen (DL) for intensity and frequency but found that "... the only factor which appeared to be related to speech intelligibility was the extent or configuration of the hearing loss." They further added that "the subjects with high-frequency hearing losses demonstrated less relative effect of noise upon their discrimination scores than did subjects with flatter pure-tone threshold configurations."

Mullins and Bangs[4] correlated some measure of the hearing levels at 500, 1,000, and 2,000 cps to PB scores and found a high correlation. Unfortunately, they did not correlate PB scores with the masking index at 1,000, 2,000, and 3,000 cps. This same study did show the correlation between PB scores and 3,000 cps to be numerically slightly greater than that between PB and 2,000 cps, and for this reason this study is quoted by Kryter as supporting the 1,000, 2,000, 3,000 cps fence.

Neither of these studies in themselves support either the 500, 1,000, and 2,000 cps or the 1,000, 2,000, and 3,000 cps fence since within any single set of data the two fences were not compared to each other.

Elliott[11] correlated PB scores to many other audiometric measures including SRT scores and the degree of the difference in hearing level at adjacent pure-tone frequencies and found that for two of her three samples "... SRT ... was the best single prediction of PB score...." In the third sample "... PB score in the non-test ear ..." was best. Although Elliott did not correlate average loss at 500, 1,000, and 2,000 cps or at 1,000, 2,000, and 3,000 cps with PB score in her original report, she has subsequently done so.[12] She finds that, "Both averages correlate negatively with PB scores ... (and) ... this negative correlation is significantly different from zero. ..." She further finds that, "Prediction of PB scores for listeners with normal hearing is not significantly different from zero for either average." In general there was no significant difference between the coefficients of PB score vs 500, 1,000, and 2,000 cps and PB score vs 1,000, 2,000, and 3,000 cps. For one of her ten groups, however, a difference between 0.66 (PB vs 500, 1,000, and 2,000) and −0.78 (PB vs 1,000, 2,000, and 3,000) was significant at the 0.05 level of confidence. There is here, therefore, a thread of evidence favoring the higher frequency fence.

Critique of Kryter's Data

The Kryter, Williams and Green[3] paper on which Kryter[2] bases his recommendations does make comparable measures. However, there is one potentially weak link in that paper[3] which throws an element of doubt into the generality of the conclusions.

The potential weak point in the Kryter, Williams, and Green[3] paper has to do with the correlation technique used in arriving at the conclusions. A correlation coefficient depends critically on the range of values studied. It is apparent from Fig 1 of Kryter et al[2] (which is reproduced as Fig 1 in Kryter[2]) that both the range of hearing losses and the number of cases with large hearing losses are less at frequencies below 2,000 cps than for frequencies above 2,000 cps. For this reason alone correlation coefficients between anything and hearing levels of 3,000 cps and above will be numerically larger than correlations of the same thing with hearing levels of 2,000 cps and below, since the magnitude of a product-moment correlation coefficient reflects to a great ex-

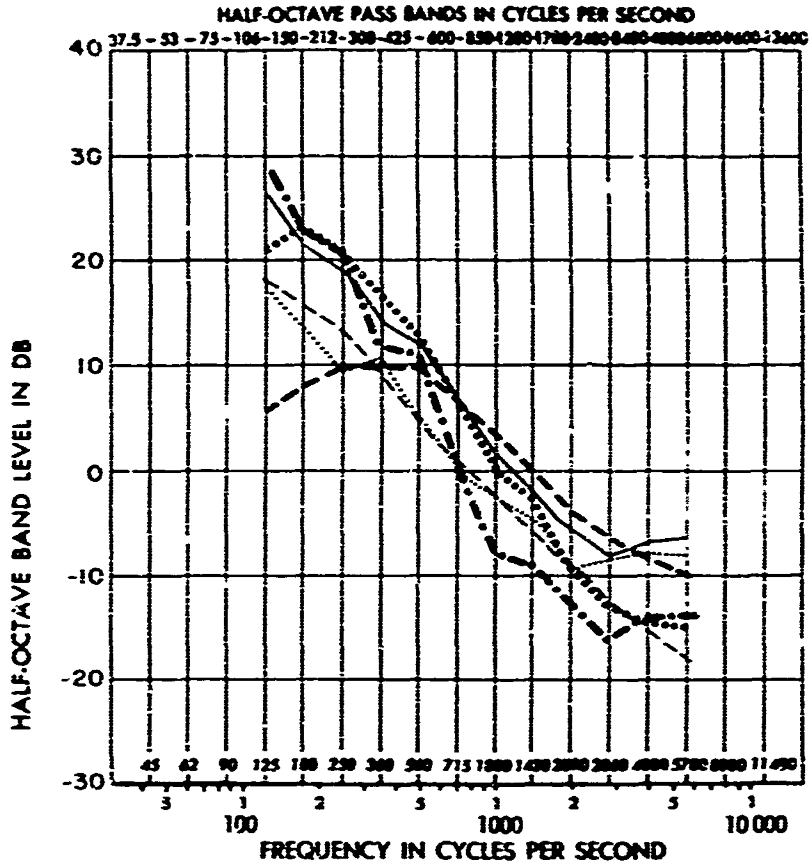


Fig 1—Masked audiograms due to low frequency noise spectra equated to be equally speech interfering with flat spectra noises.

tent the range and number of cases distributed throughout the range of measurement. This same reasoning could explain the one case out of ten in Elliott's [12] data where the 1,000, 2,000, 3,000 cps correlation with PB score was significantly greater than the corresponding 500, 1,000, and 2,000 cps correlation.

Kryter et al [3] are aware of this artifact and they point it out in comparing their data to those of Harris et al,[7] who get much higher correlations between speech threshold tests and the frequencies 500, 1,000, and 2,000 cps than the frequencies 3,000, 4,000, and 6,000 cps. Kryter et al [3] say, "These differences in results could be at least partly explained on the basis of differences among subjects . . . for example some of the subjects of Harris et al [7] suffered hearing losses that were more severe than any in our group of listeners."

If Kryter's recommendation is confined to noise-induced hearing loss subjects and if the sample he portrays (in Fig 1 of ref 2) is typical of noise-induced hearing losses, then the criticism is not particularly valid. His method, as he shows, is optimal for the sample he studied. But since the method of the AAOO's Committee on Conservation of Hearing is to encompass all types and shapes of hearing loss due to any cause, the data of Kryter et al [3] should be generalized with caution.

Although the Kryter et al [3] sample may reflect the distribution of noise-induced hearing loss, it does not necessarily represent the distribution of all kinds of hearing loss. As a general sample from which to draw general conclusions, it is not well distributed: There are essentially four groups: losses of 30 db at all frequencies (1) of 500 cps and above, (2) of 2,000 cps and above, (3) of 3,000 cps and above, and (4) of 4,000 cps and above (or out of the range of interest). There is a gaping hole in that no group qualifies with losses at 1,000 cps and above. If such a group existed, the groups would be symmetrical around the center of the dispute, namely, 1,000 and 2,000. Then averaging in either 500 or 3,000 would give some positive answer as to whether losses at 500, 1,000, and 2,000 cps or at 1,000, 2,000, and 3,000 cps would best predict hearing impairment for speech.

If Kryter's[2] basic argument is sound, that speech discrimination is more pertinent than speech threshold in determining a hearing impairment for speech, it should follow that the higher frequencies are the more important ones. But does the dropping of 500 cps for 3,000 cps really help in predicting speech discrimination loss? Note the last column of Kryter's Table I[2]: There are essentially two groups of scores, 54 and above, and 40 and below. The big break in the continuum of speech discrimination comes between those groupings. The hearing level discontinuity that divides those groups is at 2,000 cps. Both the 500, 1,000, and 2,000 cps and the 1,000, 2,000, 3,000 cps averages show large changes at this breakover point.

This same phenomenon is evident in Kryter's Fig 1.[2] To make a good bipartite prediction from graphical data it is desirable to have two horizontal lines at different levels (one corresponding to acceptable and one nonacceptable) connected in as small a horizontal spacing as possible by a steeply sloped line. The choice point is essentially the midpoint on the sloped line, and the steeper the slope the more clear-cut or precise is the prediction, or decision. On Kryter's Fig 1[2] it is evident that choosing a 5 db fence at 500, 1,000, and 2,000 cps gives exactly the same prediction as choosing a 15 db fence at 1,000, 2,000, and 3,000 cps. And, in fact, Kryter[2] makes this exact same statement when he writes, "... maintaining the 15 db fence and taking hearing level as an average of pure-tone audiograms at 1,000, 2,000, and 3,000 cps is roughly equivalent, keeping speech intelligibility constant, to measuring the hearing level at 500, 1,000, and 2,000 cps and lowering the fence by 10 db, that is, a fence of 5 db."

Based on the data he shows,[2] his recommendation could therefore have equally well been: maintain the 500, 1,000, and 2,000 cps average but lower the fence from 15 db to 5 db, which is, in fact, stated as the fourth, or last, conclusion on his paper.[2] But it should be pointed out again that no objective criterion has been mentioned, suggested, or quoted to convert Kryter's percentage of sentences—or words—correct to any location on the impairment-handicap-disability scale. Until someone, or some committee, can convert per cent PB's, sentences, or any other objective speech test to degree of impairment-handicap-disability, it is pointless to argue whether the frequencies now used to predict speech threshold from pure tone audiometer should be changed. Kryter's own data show that in actuality either set of frequencies are about equally good, depending upon what decibel level is used. Any combination of frequencies and db levels (or losses) converts only to per cent of some type of speech correct and not to any degree of impairment-handicap-disability.

A real question still remains to be answered: When does a hearing impairment for speech become a real social impairment? Beasley[12] asked people whether they considered themselves to hear normally, to be slightly impaired, etc, and then administered pure-tone audiometer tests. He found that his "minimally impaired group" averaged hearing levels of just over 20 db for 500, 1,000, and 2,000 cps and only just slightly more than that for 1,000, 2,000, and 3,000 cps. On further questioning he found his minimally impaired group had difficulty "... in church, at the theater or in group conversation, but ... not ... at close range. ..." If Beasley's population is typical of today's population, perhaps a 15 db fence is too low for a 1,000, 2,000, and 3,000 cps average, and most certainly a 5 db fence at 500, 1,000, and 2,000 cps would be too low.

If Kryter's[2] recommendation of a 15 db fence at 1,000, 2,000, and 3,000 cps be considered for adoption it should be confined to noise-induced hearing loss cases alone (which is what he recommended) and it should be kept in mind that "hearing impairment for speech" may be defined at too low a level (a man may be classified as being "hearing impaired for speech" who really, in our noisy world, is not socially "impaired" or "handicapped").

Continuing on the Beasley-type study, Dr. Leo Doerfler and Dr. Grant Fairbanks reported some preliminary works of their own at the CHABA symposium on social handicaps as related to some measure of hearing loss. This is where much effort is needed at the present time.

At this point in the CHABA symposium the present author was invited to present data showing that if speech is masked by all varieties of noises (similar in many respects to speech being heard by people with all types of audiograms) it is the noise in the octaves centered at 500, 1,000, and 2,000 cps that best describes the discrimination loss for speech. The gist of my talk follows.

Important Frequencies for Speech Perception

If consideration is given to generalizing Kryter's [2] higher-frequency hearing impairment for speech criteria to all types of hearing loss cases and in particular if a criterion for calculating hearing impairment for speech is desired that can be generalized over into the world of normal hearing people in pathological (noisy) environments, then consideration should be given to other types of evidence.

Klumpp and Webster, in a recent series of papers,[14-17] have done essentially the converse of what Kryter et al [3] have done. Whereas Kryter et al [3] gave speech discrimination tests to subjects exhibiting various limited kinds and degrees of hearing loss spectra, Klumpp and Webster [15] gave speech discrimination tests to normal hearing subjects in 16 diverse kinds of noise. In both cases subjects heard speech with part of its spectrum gone, either because of hearing loss or by being masked by noise (a rough simulation of a sensorineural hearing loss).

The Kryter et al [3] procedure is more directly applicable to hearing-impaired populations but is limited in that equal numbers of cases of all types of hearing losses were not included. The Klumpp and Webster [15] procedure is only intuitively related to hearing-impaired populations but does include a much wider distribution of masking spectra (potential simulated hearing losses of all types). Statistical analyses on the Kryter et al [3] data describe characteristics of noise-induced hearing loss as measured by speech discrimination tests. Analyses of the Klumpp and Webster [15] data describe characteristics of speech discrimination as influenced by masking noises (simulated hearing losses). There should be, and indeed there are, many similarities between the results of the two basic sets of data. Both study the discrimination of spectrum-limited speech.

In the Klumpp and Webster experiment,[15] each of eight normal-hearing subjects listened in turn to Rhyme [18] word lists in the presence of 16 diverse noise spectra. The procedure was to adjust the level of the 16 noises until each noise reduced the percentage of correctly heard words to 50%. The words were prerecorded and played back at a constant level (of 78 db at 1 meter from the loudspeaker or at the ear of the listener). The details of this experiment are described in a paper by Klumpp and Webster,[15] but suffice it to add here that Rhyme words are a series of 50 monosyllabic words like: lot, can . . . get. Each word has five alternatives (such as cot, tan, . . . let; got, fan, . . . bet, etc) so that on the answer sheets (-ot, -an, . . . -et) only the initial consonant need be supplied. Most of the 16 noises were quasi-steady-state and the spectra varied from a low-frequency rumble and generator hum, to jet aircraft noises including high-frequency whines. There was a typewriter sound and a multivoice babble. Laboratory-generated thermal noises of flat spectrum and of an increasing 6 db per octave spectrum (predominantly high frequency) and of a decreasing 6 db per octave were also used. These noises, unlike clinical hearing levels, simulated as many low-frequency or high-frequency losses as flat losses.

The purpose of the original papers [14,15] was to find simple means of measuring the

16 noises which had already been adjusted in level to be equally speech-interfering. The analogous problem in clinical audiology would be to find the common hearing level pattern of a group of cases all of whom scored 50% on a word list presented at the same level to each of them.

In the first published version of these results,[14] it was found that the best simple measure of predicting the speech interference of these noises was to find the level of noise, measured in db, in the octaves 300 to 600, 600 to 1,200, and 1,200 to 2,400 cps and then average these three levels (roughly analogous to finding the average hearing level at 425, 850, and 1,700 cps). A subsequent paper [15] found that an average of the noise levels, in db, of octaves centered at 500, 1,000, and 2,000 cps was nearly equivalent to the slightly lower octaves of 425, 850, and 1,700 cps.[14] In both cases these results were better than average levels for octaves centered at 850, 1,700, and 3,400; or at 1,000, 2,000, and 4,000 cps.

Likewise, it was found that the one single octave that predicted better than any other single octave was the octave centered at 850 (or 1,000) cps, followed in order by 425 (or 500) cps, 1,700 (or 2,000) cps, and finally 3,400 (or 4,000) cps.

To the extent that the level of a masking noise in a given region corresponds to a hearing loss in the same region, the important frequencies are in order: 1,000, 500, 2,000, and 4,000 cps, and the average of 500, 1,000, and 2,000 cps is superior to the average of 1,000, 2,000, and 4,000 cps.

The Importance Frequency in Speech Perception

The center, mid, or mean frequency of Kryter's[2] 1,000, 2,000, 3,000 cps scheme is 1,817 cps (cube root of the product) whereas the mean frequency of the Committee on Conservation of Hearing scheme is 1,000 cps. If this midfrequency be called the most important frequency in speech perception, there is a large amount of literature on the subject.

French and Steinberg,[5] using speech in the quiet and normal listeners, progressively high- and later low-pass-filtered the speech until it became progressively less intelligible. This was a discrimination, not a threshold test. They found that speech was equally deteriorated when all frequencies either above or below 1,900 cps were filtered out. Or the frequency range above 1,900 cps was as important as the frequency range below 1,900 cps. Beranek,[19] using male voices only, found the crossover frequency to be 1,660 cps under the same quiet-filtered-speech conditions.

Pollack[20] redid the filtered speech intelligibility studies but added a broad-band noise background and varied the level of the speech. He found that the crossover (or equal importance) frequency increased from 800 cps for low levels of speech through 1,010, 1,300, 1,430, to 1,620 cps for increases of 10 db in the speech level.

Dyer,[21] doing the reverse of Pollack (namely, filtering the noise around a broad band speech signal), found, like Pollack, that as the speech-to-noise differential increased, the crossover frequency increased from about 1,000 cps to almost 2,000 cps.

A crossover frequency has been derived from the Klumpp and Webster[15] data and published along with details of its derivation.[16] The data in condensed form are shown in Fig 1-4. Fig 1 shows the masked audiograms due to the predominantly low frequency noises from Webster and Klumpp.[16] The masked audiograms are plotted as differences from the masked audiogram due to the thermal or flat noise, but the relative levels and slopes are representative of the noises as adjusted to be equally speech-interfering. Fig 2 and 3 show similar plots (again related to the flat noise masked audiogram) of the slightly low frequency and nearly flat spectra noises. The area enclosed by the spread of the noise audiograms in Fig 1, 2, and 3 is plotted in Fig 4 at the top. At the bottom of Fig 4 are the average masked audiograms of Fig 1, 2, and 3 together with the one predominantly high frequency noise-masked audiogram. Observe in

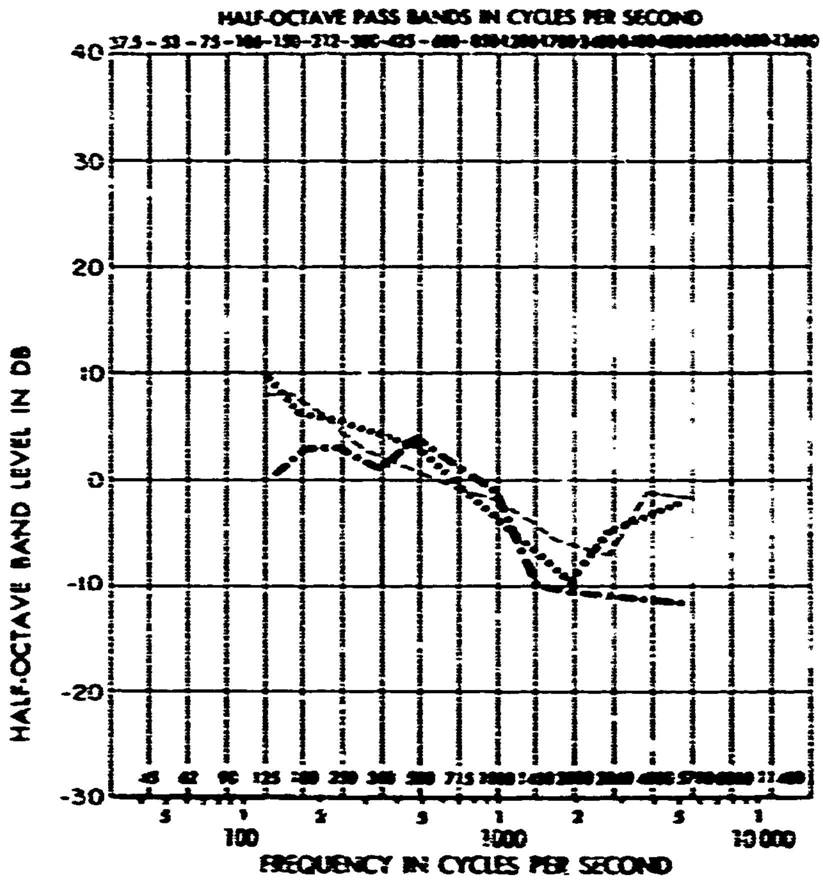


Fig 2—Masked audiograms due to equally interfering semi-low frequency noises equated to flat thermal noise.

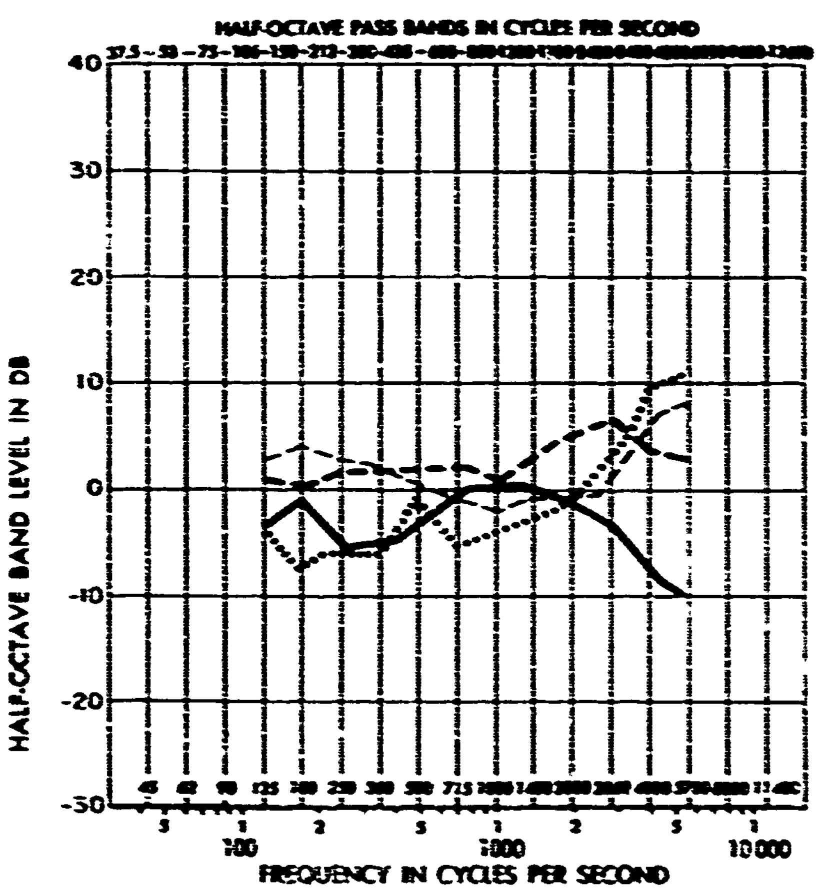


Fig 3—Masked audiograms due to equally speech interfering flat spectra noises, ie, noises generally equivalent to thermal noises.

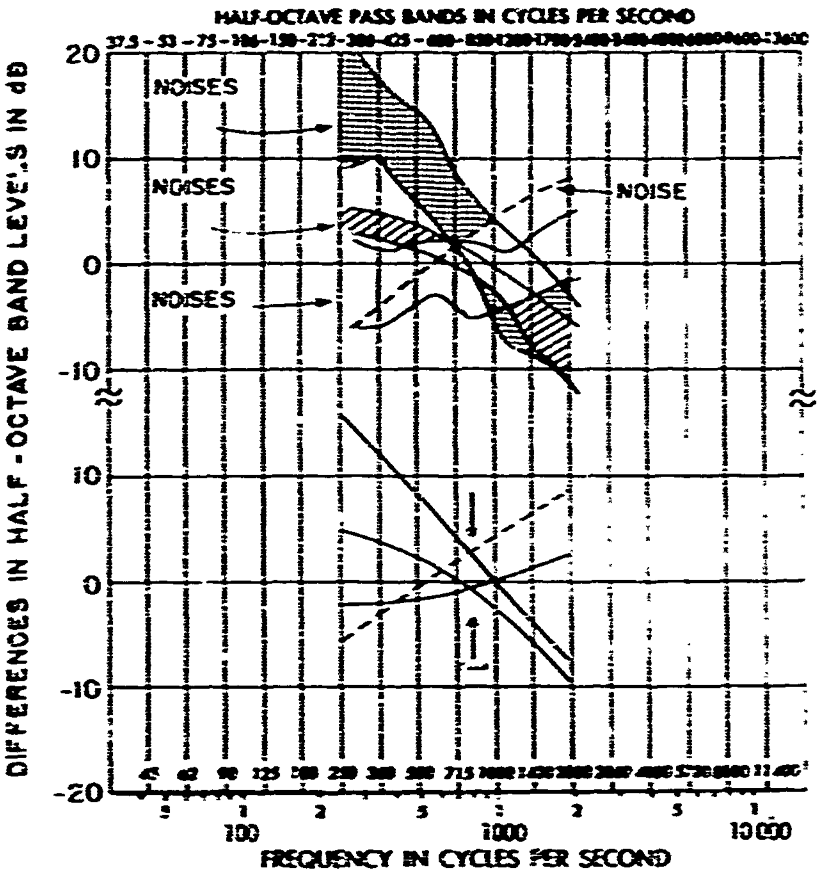


Fig 4.—Superposition of masked audiograms due to spectra of equally speech interfering noises, showing crossover or importance frequency.

Fig 4 that both the predominantly high- and low-frequency noise audiograms and the slightly high- and low-frequency noise audiograms cross each other at about 850 cps.

Kryter[22] also has some data of speech masked by noise which show a low crossover frequency. If the information in his Fig 2 and 3 of ref 22 were recombined to show the levels of the noise spectra (in Fig 2, ref 22) plotted to yield equally speech-interfering scores of 60% (from Fig 3, ref 22) the two steeply sloped noises (predominantly high- and low-frequency noises) would cross at about 900 cps and the more gently sloped high- and low-frequency noises at 1,400 cps.

It appears therefore from the evidence of Pollack,[29] Dyer,[23] and Kryter,[22] and Webster and Klumpp[16] that noise-masked speech has a crossover or importance frequency as much as an octave lower than the crossover frequencies of filtered speech in the quiet (French and Steinberg,[5] and Beranek[19]). Both Pollack[29] and Dyer[23] show that the frequency varies from 800 or 1,000 cps to 1,600 or 2,000 cps as the speech-to-noise differential increases.

These apparent anomalies led to further work on speech interference criteria. In a paper by Webster,[17] the articulation index (AI) developed by French and Steinberg[5] and its application in noises of various levels led to a set of contours which seems to explain what happens when listening to speech in increasing levels of noise (or increasing hearing losses). Fig 5 shows that as the level of noise in an octave band increases (analogous to increasing hearing losses), the minimum point of the contour that traces out any one given articulation index (AI) level decreases in frequency. In simple terms, as the listening conditions get worse (more noise or presumably, greater hearing losses), the importance frequency gets lower.

In terms of the original argument of whether 500, 1,000, and 2,000 cps midfrequency 1,000 cps; or 1,000, 2,000, and 3,000 cps, midfrequency 1,817 cps, be used as a

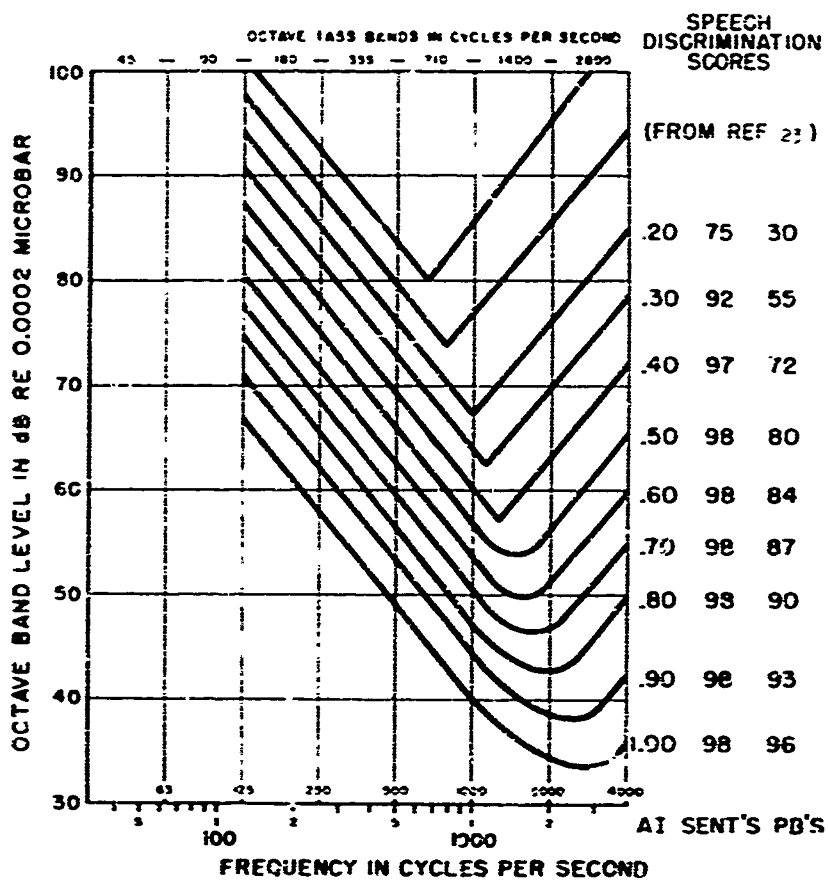


Fig 5—Permissible noise levels for indicated Articulation Index (AI) scores when speech is spoken at a "conversational level." AI contours and the corresponding sentence and PB word scores (taken from ref 23) are indicated at the right.

basis for predicting hearing impairment for speech, the answer seems to depend on what level be chosen as "impaired." If very little impairment will be tolerated, say, an AI of 0.60 (sentence intelligibility of 98%), the most efficient method will be an average of some higher frequencies, say, 1,000, 2,000, and 3,000 cps. If a greater degree of impairment will be tolerated, say, an AI of 0.30 (sentence intelligibility of 92%), an average frequency centered just above 1,000 cps would be most efficient.

To make the pure-tone audiogram average used to predict hearing impairment for speech as universal as possible (to cover hearing loss and to be compatible with noise masked speech), it is suggested that Kryter's[2] proposed 1,000, 2,000, and 3,000 cps fence not be adopted until further evidence is available. The further evidence should definitely include a more precise delineation of when does a "speech discrimination loss" become a "hearing impairment for speech."

Summary of the Klumpp and Webster Data

The Klumpp and Webster[15] data show that when speech is masked by noise to the extent that only 50% of initial consonants of single words are heard correctly, the average of the levels of noise in the octaves centered at 500, 1,000, and 2,000 cps predicts the masking effect of the noise better than averages in any other groupings of octaves.

The Webster paper[17] shows that when speech is masked by noise, the frequency that divides the speech into two band-widths, each of which contributes equally to the intelligibility of the speech, changes as the speech-to-noise ratio (S/N) changes. At low levels of S/N the dividing (crossover, or importance) frequency is around 1,000 cps. As the S/N increases, the importance frequency increases to about 2,000 cps.

Conclusion

I believe that no decision can be made as to what is the most efficient combination of pure-tone losses to predict ability to hear speech until a point is found that defines some "compensable," or otherwise adequately defined "handicap" in the impairment-handicap-disability scale. If the criterion point is close to the "impairment" end of the scale, then perhaps a center frequency around 1,800 cps is the best predictor. If, however, the criterion point is defined more in the center or toward the disability end of the scale, a center frequency at or below 1,000 cps would be the best predictor.

John C. Webster, PhD, US Navy Electronics Laboratory, San Diego, Calif 92152.

REFERENCES

1. Lierle, D. M.: Report of the Committee on Conservation of Hearing, Trans Amer Acad Ophthal Otolaryng 63:236 (March-April) 1959.

2. Kryter, K. D.: Hearing Impairment for Speech: Evaluation From Pure-Tone Audiometry, Arch Otolaryng 77:598-602, 1963.

3. Kryter, K. D.; Williams, C.; and Green, D. M.: Auditory Acuity and Perception of Speech, J Acoust Soc Amer 34:1217-1223, 1962.

4. Mullins, C. J., and Bangs, J. L.: Relationships Between Speech Discrimination and Other Audiometric Data, Laryngoscope 67:1-15, 1957.

5. French, N. R., and Steinberg, J. C.: Factors Governing the Intelligibility of Speech Sounds, J Acoust Soc Amer 19:90-119, 1947.

6. Quiggle, R. R., et al: Predicting Hearing Loss for Speech From Pure Tone Audiograms, Laryngoscope 67:1-15, 1957.

7. Harris, J. D.; Haines, H. L.; and Meyers, C. K.: New Formula for Using the Audiogram to Predict Speech Hearing Loss, AMA Arch Otolaryng 63:158-176, 1956.

8. Hanley, C. N.: Factorial Analysis of Speech Perception, J Speech Hearing Dis 21:76-87, 1956.

9. Solomon, L. N.; Webster, J. C.; and Curtis, J. F.: Factorial Analysis of Speech Perception, J Speech Hearing Res 3:101-107, 1960.

10. Ross, M., et al: Speech Discrimination of Hearing Impaired Individuals in Noise, USAF School of Aerospace Medical Technology Documents Report No. 62-130, April, 1963.

11. Elliott, L. L.: Prediction of Speech Discrimination Scores From Other Test Information, USAF School of Aerospace Medicine Technology Documentary Report No. SAM-TDR-62-145, Dec, 1962; also published in J Aud Res 3:35-45, 1963.

12. Elliott, L. L.: Personal communication to the author, May 11, 1964 (major contents of which will be submitted as a Letter to the Editor in J Acoust Soc Amer).

13. Beasley, W. C.: Characteristics and Distribution of Impaired Hearing in Population of the United States, J Acoust Soc Amer 12:114-121, 1940.

14. Klumpp, R. G., and Webster, J. C.: Predicting Speech Interference From Physical and Psychophysical Measures of Ambient Noise, Proceedings of the Speech Communication Seminar, Stockholm, Sept, 1962.

15. Klumpp, R. G., and Webster, J. C.: Physical Measurements on Equally Speech-Interfering Navy Noises, J Acoust Soc Amer 35:1328-1338, 1963.

16. Webster, J. C., and Klumpp, R. G.: Articulation Index Validation on 16 Equally Speech-Interfering Noises, J Acoust Soc Amer 35:1339-1334, 1963.

17. Webster, J. C.: Generalized Speech Interference Noise Contours, J Speech Hearing Res 7:133-140, 1964.

18. Fairbanks, G.: Test of Phonemic Differentiation: Rhyme Test, J Acoust Soc Amer 30:596-600, 1958.

19. Beranek, L. L.: Design of Speech Communication Systems, Proc IRE 35:880-890, 1947.

20. Pollack, I.: Effects of High Pass and Low Pass Filtering on Intelligibility of Speech in Noise, J Acoust Soc Amer 20:259-266, 1948.

21. Dyer, W. R.: Masking of Speech by High- and Low-Pass Noise, Technical Document Report No. RADC-TDR-298, Rome Air Development Center, July, 1962.

22. Kryter, K. D.: Validation of Articulation Index, J Acoust Soc Amer 34:1698-1702, 1962.

23. Hawley, M. E., and Kryter, K. D.: "Effects of Noise on Speech" in Handbook of Noise Control, edited by C. M. Harris, New York: McGraw-Hill Book Co., Inc., 1957, chap 9, Fig 9-5.

Reprinted from THE JOURNAL OF ACOUSTICAL SOCIETY OF AMERICA, Vol. 37, No. 4, 692-699, April 1965

Speech Communications as Limited by Ambient Noise

J. C. WEBSTER

U. S. Navy Electronics Laboratory, San Diego, California
(Received 17 August 1964)

Speech-intelligibility scores as a function of noise level are studied for face-to-face, sound-powered-phone, and amplified speech- (earphone and loudspeaker) communication conditions. The speech-interference level (SIL) for octaves of noise centered at 500, 1000, and 2000 cps (0.5/1/2) is used as the measure of noise level. By using this noise measure, much of the work in this field can be brought together and interpreted. It is noted that "noisy" and "very noisy" spaces are associated with SIL's such that "shouting" or "very loud" voice levels (or 95-dB speech levels) are required for conversations at 1.5 or 3 ft, and this is the region where telephone conversations are judged to be "difficult" or "unsatisfactory." All of these adverse noise conditions occur at the region where ear protection will aid intelligibility and at the boundary where ear protection should be used to protect against hearing losses. Where people must converse or communicate via some interior communication device, 0.5/1/2 SIL's above 70 dB should be avoided. At 0.5/1/2 SIL's greater than 90 dB, the wearing of hearing protection should be made mandatory and every noiseproofing technique (except a noise shield for the microphone) should be employed. At 0.5/1/2 SIL's above 100 dB, every noiseproofing technique should be employed.

INTRODUCTION

OVER the past few years, several aspects of speech intelligibility in noise have been studied at the U. S. Navy Electronics Laboratory (NEL). Many speech-communication equipments and/or systems have been evaluated for operation both in the quiet and in noise. Many of these evaluations have come out as NEL reports[1-7] or were published in journals not readily available.[8,9] Others, especially those dealing with the noise-attenuating properties of ear protection,[10,11] have appeared in available journals but the application of the results to communication systems need be examined. In many of these evaluations, the results were comparative in nature (one system or equipment was evaluated in terms of another) and no attempt was made to interpret the results in absolute terms.

Relating the results of these evaluations to each other and to the work of others has been difficult because of the diverse ambient noises used and the differences in the way both speech levels and noise levels have been measured. A recent series of papers[12-15] concerning methods of predicting speech interference provides a framework that makes it easier to generalize some of the evaluative results. It is the purpose of this paper to

[1] J. C. Webster and R. G. Klumpp, "USNEL Flight Deck Communications System. Part 2. Noise and Acoustic Aspects," NEL Rept. 923 (29 Nov. 1960), AD 260 286.

[2] W. E. Montague, "A Comparison of Five Intelligibility Tests for Voice Communication Systems," NEL Rept. 977 (27 June 1960), AD 254 545.

[3] J. C. Webster and R. G. Klumpp, "Evaluation of the AN/PRC-53," NEL Rept. 1042 (18 Apr. 1961), AD 260 294.

[4] J. C. Webster, P. O. Thompson, and T. H. Wells, "Evaluation of the AN/PRC-44(XN-1)," NEL Rept. 1058 (30 July 1961).

[5] J. C. Webster and P. O. Thompson, "Noise-Proofed Sound Powered Phones," NEL Rept. 1073 (25 Oct. 1961).

[6] J. C. Webster, P. O. Thompson, and H. R. Beitscher, "Intelligibility of Amplified Speech in Helicopter Noise," NEL Rept. 1080 (7 Nov. 1961).

[7] J. C. Webster and R. G. Klumpp, "Technical Evaluation of the AN/SRC-22 (XN-1) Flight Deck Communication System," NEL Rept. 1141 (15 Oct. 1962).

[8] J. C. Webster and P. O. Thompson, "Dynamic or Carbon Microphone?" Bur. Ships J. 10, 8-9 (14 Sept. 1961).

[9] J. C. Webster, R. G. Klumpp, and P. O. Thompson, "Capabilities of Speech Communication in Noise," Proc. Inst. Environ. Sci., 297-307 (Apr. 1962).

[10] J. C. Webster, "Ear Defenders: Measurement Methods and Comparative Results," Noise Control 1, No. 5, 34-42 (1955).

[11] J. C. Webster and E. R. Rubin, "The Noise Attenuation of Selective Ear Protective Devices," Sound—Its Uses and Control 1, No. 5, 34-46 (1962).

[12] R. G. Klumpp and J. C. Webster, J. Acoust. Soc. Am. 35, 1328-1338 (1963).

[13] J. C. Webster and R. G. Klumpp, J. Acoust. Soc. Am. 35, 1339-1344 (1963).

[14] R. G. Klumpp and J. L. Leonard, "Observer Variability in Reading Noise Levels with Meters," Sound—Its Uses and Control 2, No. 4, 25-29 (1963).

[15] J. C. Webster, "A Speech Interference Noise-Rating Contour," Acustica (to be published).

NEL | A CONTRIBUTION OF THE NAVY ELECTRONICS LABORATORY, SAN DIEGO, CALIFORNIA 92152

summarize capabilities of speech communication in noise and in some instances to compare the results with other studies not directly concerned with equipment evaluation.

I. OVERVIEW OF SPEECH INTELLIGIBILITY IN NOISE

The Rosetta stone that allows a comparison among the NEL evaluations and between them and other studies in the general field of speech intelligibility in noise is shown in Table I. Table I is the Rosetta stone if it is assumed that it is the level and spectral distribution of noise that most affect speech intelligibility. Factors such as the choice of speech materials, and the selection and training of talkers and listeners, preclude making exact comparisons.

Table I (and its predecessors, Table II and Fig. 18 of Ref. 12, and Table I and Fig. 3 of Ref. 15) shows physical measures, such as an A, a C, and a proposed SI (speech interference)[15,16] sound-level meter reading, and speech-interference level (SIL) measures[17] (at 600–4800, 300–4800, and 500, 1000, 2000 cps)[18] for noises of different spectra that are equated to be equally speech-interfering. In addition to listing the levels of different noises that are equally speech-interfering, Table I also lists several equivalent measures of any single noise. If the spectrum of an unknown noise resembles one of the noises in Table I, any single measure of this noise can be used to find other measures and thereby facilitate comparing the results of one study or evaluation with others.[19]

With this much explanation of Table I, it is time to look at the general overview of the various NEL evaluation results. Figure 1, a generalized summary of many existing studies,[1,3–5,7,9] shows the scope of the problem of talking and listening in noise. On the ordinate is plotted percentage of rhyme words[20] heard correctly and along the abscissa is the level of noise.

TABLE I. Levels in dB of noises producing 50% scores for rhyme words.

No.	Noise Name	Sound-level meter C	A	SI-70[a]	Speech-interference levels 0.6/4.8[c]	0.3/4.8[c]	0.5/1/2[d]
1	Rumble (−12)	105	86	83	66	71	72
3	Blower (−9)	92	85	85	67	72	74
4	TN −6	87	79	81	69	72	73
...	Helicopter[b]	(82)	(82)	(80)	(73)	(72)	(74)
9	Babble	81	82	80	75	73	73
10	TN flat	81	82	80	75	73	73
15	TN +6	88	90	84	80	77	75
16	Jet	94	94	86	81	80	79
	Average	88.8	85.0	82.4	73.2	73.8	74.1
	Range	24	15	6	15	9	7

[a] Calculated (not measured) and equated with "A" on noise No. 3.
[b] Not in original (Ref. 12) study, equated on "SI-70" with noise No. 10.
[c] These are band-limiting frequencies; the band center frequencies are 0.85/1.7/3.4 and 0.43/0.85/1.7.
[d] These are the center frequencies of octave bands; the band-limiting frequencies are 0.35/2.8.

The choice of the 0.5/1/2 SIL of an equivalent −6-dB/oct thermal noise is based on two facts. The first is that the 0.5/1/2 SIL is a reasonable compromise for showing small numerical fluctuations among the physical measurements of the original 16 equally

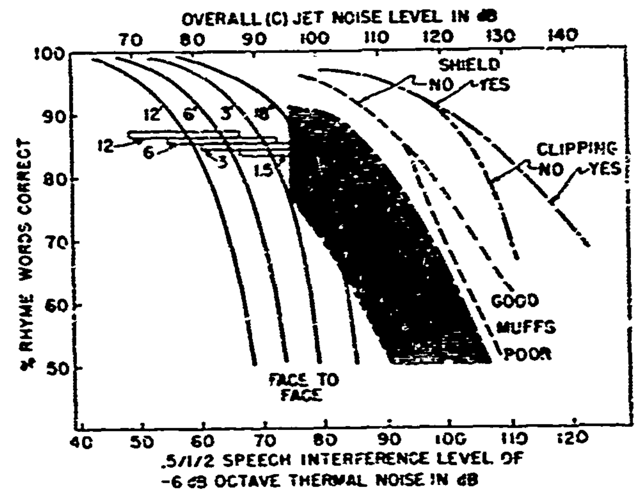


FIG. 1. Speech intelligibility (percent rhyme words correct) as a function of jet-aircraft idling noise level. On the top abscissa, noise levels are listed as measured on the C-weighting network of a sound-level meter. On the bottom abscissa, the noise level is listed as the speech interference level (SIL), based on the octaves centered at 500, 1000, and 2000 cps, of a minus 6-dB/oct shaped thermal noise that is equivalent in its ability to interfere with the intelligibility of speech to the jet-noise levels listed on the top abscissa. Three generic types of results are shown: face-to-face, sound-powered phone, and amplified speech. Within the face-to-face results, the parameter is distance between talker and listener. The limits on the sound-powered results are present-day "operational" equipment (to the left) and "developmental" equipment (to the right). In the amplified-speech results, the major parameter is presence or absence of a microphone shield. When a shield is used, a subparameter is whether or not clipping is used for earphone listening. When a shield is not used, the subparameter is whether an average (left) or excellent (right) earmuff is used around the earphone.

[16] J. C. Webster, J. Acoust. Soc. Am. 36, 1662–1669 (1964). The frequency-dependent curve upon which the SI filter is based is shown in Fig. 2, Ref. 15, or as the SI-70 curve in Fig. 1 of this reference. To make the filter, the inverse of this SI-70 curve is used. The 70 refers to the fact that origional curve was based on noises adjusted in level to be equally speech-interfering at an SIL level of just over 70 dB (see Table I in this paper or Table II, Ref. 12).

[17] L. L. Beranek, "Airplane Quieting II—Specifications of Acceptable Noise Levels," Trans. ASME 69, 97–100 (1947).

[18] SIL's are based on arithmetic averages of noise levels in octave bands. The octave bands can be specified by lower and upper band limits, or by center frequencies. The band centers of octaves ranging from 600 to 4800 cps are 850, 1700, and 3400. The band limits of octaves centered at 500, 1000, and 2000 cps are approximately 350 to 2800 cps. Because of past (band [illegible]) usage, 0.6/4.8 SIL (not 8.5/17/34) is used to specify the SIL calculated from the octaves 600–1200, 1200–2400, 2400–4800 cps. And to maintain the simplicity inherent in its choice, 0.5/1/2 SIL (not 0.35/2.8) is used for the SIL for octave bands centered at 500, 1000, and 2000 cps.

[19] R. W. Young, J. Acoust. Soc. Am. 36, 289–295 (1964). Young has compiled a Table similar to Table I, associating various measures of noise to idealized noise spectra. Young's Table is not for equally speech-interfering noises, but for noises equal on A weighting, which for existing sound-level-meter networks is the best for predicting speech interference.

[20] G. Fairbanks, J. Acoust. Soc. Am. 30, 596–600 (1958).

speech-interfering noises.[12] The second is that the −6-dB/oct noise is a reasonable compromise among noises to be representative of ship noises,[12] office noises,[21,22] and noises used in laboratory studies of speech intelligibility.[23]

Figure 1 deals with three specific communication situations: face-to-face, sound-powered phone, and amplified speech. It shows for each form of communication the limiting noise levels for given degrees of communication effectiveness (percentage of rhyme words correct). For the majority of the studies summarized in Fig. 1, a single experienced talker and five experienced listeners were used.[9] The following three sections deal in detail with each communicating situation: face-to-face, sound-powered phone, and amplified speech. In each section, repeated reference is made back to Fig. 1.

II. FACE-TO-FACE COMMUNICATION

The most satisfactory, but least noise-resistant, communication is face-to-face communication. For these face-to-face tests, no constraints were put on the vocal level of the talker. Nor was the talker asked to maintain any given level of word intelligibility. He was in the same room with his listeners. He could see them and he could hear the ambient-noise level around them. The listeners faced the talker but the rate of word presentation was such that lip reading can be almost completely discounted. The voice level that he adopted was left to his own knowledge of the test situation.

The limiting factor in face-to-face communication in noise is the distance between the talker and the listener, since the potential voice level of the talker and acceptable listening levels are physiologically limited. Observe in the four curves to the left in Fig. 1 that at any single criterion level, say 70% correct, for each doubling of the distance between talker and listener, 6 dB less noise can be tolerated.

To show how the face-to-face data in Fig. 1 compare with the extensive work of Beranek,[17,21,22] which culminated in the 0.6/4.8 SIL, note the bars superimposed at the 85% rhyme-word position between the 0.5/1/2 SIL levels of 48 and 84 dB. For ease of viewing, the bars are placed under one another; however, they should be thought of as being located at, and only at, the 85% point. The 85% position was chosen because Ref. 24 states that barely reliable conversation corresponds to a PB word score of 75%. And Montague[2] shows that a PB score of 75% corresponds to a rhyme score of 85% (and a Harvard sentence[25] score of 93%).

The SIL bars represent, from top to bottom, the range of noise levels in which conversation can be carried on when the talkers and listeners are 12, 6, 3, and 1.5 ft apart. The left end of each bar represents "normal" voice level, the right "shout"; equally spaced in between are "raised" and "very loud" voice levels. It is understood, of course, that if a "normal" level is understandable at any given level of noise it will also be understood at lesser levels of noise.

The SIL bars were placed between 48 and 84 dB by adding 5 dB onto the levels shown in any Table showing 0.6/4.8 SIL versus distance and voice level (say, Table 4-13 of Ref. 24). This 5 dB accounts for the difference between the 0.6/4.8 SIL and 0.5/1/2 SIL's for noises similar to noises 3 and 4 in Table I. Noises 3 and 4 are typical of the office noises used by Beranek[17,21,22] in developing the SIL.

The fact that the bars (representing Beranek's SIL formulation) are almost exactly bisected by the proper distance contours (where voice level was at the option of the talker) shows that the two sets of data are highly compatible.

The maximum noise level in which face-to-face communication is possible can probably be set at about 0.5/1/2 SIL of 95 dB. Note, for example, the maximum rating in 0.6/4.8 SIL Tables[24] is 89 dB (or 94 dB 0.5/1/2 SIL). Or note that Pickett[26] states, "The maximum tolerable noise levels for 90% sentence intelligibility and 1 m between talker and listener were estimated to be 95 dB for white noise [noise No. 10, Table I] and 105 dB for low-frequency noise." These over-all levels, when converted (from the spectra given in Pickett's original paper) to 0.5/1/2 SIL, become 88 and 96 dB, respectively. Pickett's[26] conditions were for a 1-m distance between talker and listener whereas the 0.6/4.8 SIL Table specified 0.5 ft. But Pickett was using a highly trained talker and looking specifically for a maximum level. The 0.6/4.8 SIL Table is for average talkers and listeners.

An extrapolation to Fig. 2 of Ref. 27 also shows an over-all level of 95-dB thermal noise (0.5/1/2 SIL = 87 dB) to evoke the maximum usable speech level from talkers communicating at 1 m. Face-to-face communications were not attempted in levels this high in the data of Fig. 1. An extrapolation of the "distance-halved" contours at 85% rhyme scores at 95 dB 0.5/1/2 SIL shows the permissible distance to be roughly 2.25 in. All of these values tend to confirm the fact that face-to-face (actually mouth to hearing-protected ear) communication ceases at about 115 dB(C) in jet noises on aircraft carriers.

[21] L. L. Beranek, J. Acoust. Soc. Am. 28, 833–852 (1956).
[22] L. L. Beranek, "Revised Criteria for Noise in Buildings," Noise Control 3, No. 1, 19–27 (1957).
[23] K. D. Kryter, J. Acoust. Soc. Am. 18, 413–417 (1946).
[24] K. D. Kryter, J. C. R. Licklider, J. C. Webster, and M. Hawley, "Speech Communication," in *Human Engineering Guide to Equipment Design*, C. T. Morgan *et al.*, Eds. (McGraw-Hill Book Co., Inc., New York, 1963), Chap. 4, p. 179.
[25] J. P. Egan, "Articulation Testing Methods," OSRD Rept. 3802 (1944); PB 22848. Also published in Laryngoscope 58, 955–991 (1948).
[26] J. M. Pickett, J. Acoust. Soc. Am. 30, 278–281 (1958).
[27] J. C. Webster and R. G. Klumpp, J. Acoust. Soc. Am. 34, 936–941 (1962).

III. SOUND-POWERED-PHONE COMMUNICATION

The sound-powered phone (SPP), as its name implies, requires no source of external power. Its sole source of audiofrequency energy is sound, preferably voice sounds but also, of course, noise. It achieves an orthotelephonic gain of 20 dB (a gain of 20 dB over the level of a voice in air at 1 m) by utilizing a resonance phenomenon[25] to take full advantage of the existing energy in a narrow band around 1500 cps. It sacrifices quality (wide-band frequency response) for quantity (more speech power per unit bandwidth). Typical frequency responses of sound-powered phones[29] show peaks at frequencies varying from 800 to 2500 cps, and very little response at all beyond plus or minus 1 oct from the peak frequency.

Wiener[25] states that "The rapid decrease in response above 1500 cps is due to both earphone and microphone. . . . The decrease . . . at the lower frequencies is due mainly to the microphone." In discussing whether improved performance could be obtained by centering the resonance at another (lower) frequency, Wiener (based on data of Egan and Wiener, Ref. 28, Sec. 7.2.4, but published elsewhere later[30]) concludes, " . . . net gain in performance will result from shift in frequency . . . only if the orthotelephonic gain or the bandwidth or both are increased." Since SPP communication is inherently limited by its frequency-response characteristics, SPP intelligibility scores for quiet conditions are typically between 85% and 90% PB words correct rather than between 95% and 100% as are scores for broad-band equipment. Because of this, Wiener[25] questioned (in 1946) " . . . whether it might not be more profitable to abandon the sound-powered principle . . . and install an interphone system using efficient wide-band instruments and electronic amplifiers." In refuting his own doubts, Wiener[25] listed reasons for the continued widespread use of sound-powered phones in Naval ships. He mentioned "reliability" and the fact that the sound-powered phone " . . . performs satisfactorily under conditions of relative quiet . . . [and] . . . is the simplest [system] possible." He did, however, suggest, and made preliminary tests to show, the advantages of noise-cancelling microphones and noise shields for microphones, and the use of better ear cushions around the earphones. All of these schemes have subsequently been tested and have extended the potential use of sound-powered phones into higher and higher noise fields.

Whether the SPP system should remain the mainstay of USN ships' interior communications in this day of higher noise levels (and increasing needs for passing more complex data and information) is a systems problem beyond the scope of this paper. The system aspects of the problem are covered elsewhere.[31] The purpose of this section of this paper is to assess the range of noise levels in which speech communication is possible, using sound-powered phones.

When the same five listeners and the same talker used in the face-to-face experiment are tested on SPP equipment, the SPP results summarized in the center of Fig. 1 obtain. With present-day operational, nonnoise-proofed phones, the results are no better than face-to-face communication at, say, 2 ft (considering the 70% criterion). However, "developmental" equipment utilizing noise-cancelling microphones and noise-attenuating cushions around earphones does extend usable communications to noise levels beyond face-to-face capabilities.

One of the differences between face-to-face communication and a wire-connected system is that in a wire-connected system each communicator can be (and usually is) in his own noise environment. Figure 1 is based on the premise that both the talker and the listener are in the same noise environment. Figure 2 reflects speech-intelligibility results when there are different noise environments around the talker and the listener.[32] Three possible cases are shown: N-N, N-Q, and Q-N. In the N-N case, both the talker and listener are in the noise level shown on the abscissa. These are the data summarized in Fig. 1. In the N-Q case, the talker is in the noise level along the abscissa and the listener is in relative quiet (the noise level of the leftmost

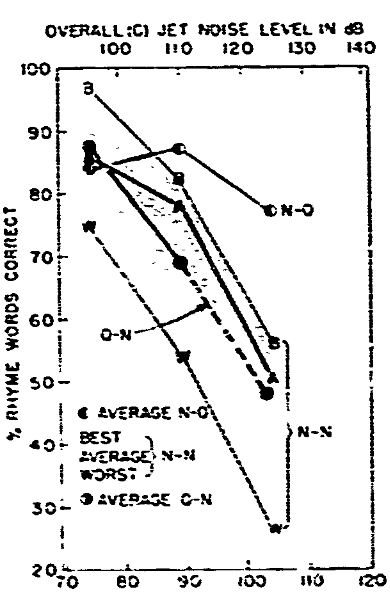


FIG. 2. Speech intelligibility (percent rhyme words correct) vs jet-noise level for SPP equipments. When both the talker and the listeners are in the noise levels listed on the abscissa (N-N), B refers to developmental equipment, W to operational equipment, and A to an average of about 8 equipments (mostly developmental). For the N-Q and Q-N conditions, only the average results are shown. N-Q means talker in noise, listeners in quiet (noise level of leftmost data point). Q-N is the reverse: talker in "quiet," listeners in noise. The noise levels are as in Fig. 1: over-all (C) level of jet noise (top) and 0.5 1, 2 SIL of a −6-db oct noise that is equivalent to the jet noise in its speech-interfering properties.

[25] F. M. Wiener, "Special Voice Communication Systems," in *Transmission and Reception of Sounds under Combat Conditions*, C. E. Waring, Ed. (Sum. Tech. Rept. Div. 17 National Defense Research Commission, Washington, D. C., 1946), Vol. 3, Chap. 12.

[29] L. L. Beranek (and staff of Electro-Acoustic Lab., Harvard Univ.), "Audio Characteristics of Communication Equipment," PNR-6 (1 Feb. 1945), Figs. 46-50.

[30] J. P. Egan and F. M. Wiener, J. Acoust. Soc. Am. 18, 435-441 (1946).

[31] J. C. Webster and I. B. Henry, "Do Sound-Powered Phones Have a Future?" (to be published).

[32] The data are from Ref. 5 and P. O. Thompson, "Effectiveness of Three Sound Powered Telephone Sets in High-Level Noise," NEL Tech. Mem. 607 (11 June 1963).

data point in each series). In the Q-N case, the talker is in the noise level of the leftmost data point and the listener is in the noise level shown on the abscissa.

There is, of course, the fourth case (Q-Q) where both the talker and listener are in relative quiet. This is the limiting case for each of the other three cases. In Fig. 2, the spread of the leftmost points of each of the average conditions (85±2%) is a measure of the variability of the experimental data. Note in Fig. 2 that, when averaged over many different (operational and developmental) sound-powered phones, the highest intelligibility scores obtain in the N-Q case, then the N-N case, and the lowest scores are for the Q-N case (since the talker doesn't "speak up" when in the quiet). The "BEST" results almost always obtain when the latest and best developmental "noiseproofed" equipment is tested. The "WORST" intelligibility results are from the older operational equipment not designed for use in high noise levels.

Although all of the data are not duplicated here from Refs. 5, 9, 32, and 33, some generalizations from the more detailed data are made: namely, the best developmental noiseproofed sound-powered equipment can be used (70% rhyme words correct) in noise levels of 114 dB 0.5/1/2 SIL of −6-dB/oct thermal noise (TN-6) or 135 dB over-all jet noise *if* the listener is in a quiet location (N-Q); 94 dB SIL or 115-dB over-all jet noise when both talker and listener are in noise (N-N); and 84 dB SIL or 105-dB jet noise if the talker is in the quiet and cannot be induced to literally shout into his microphone (Q-N).

IV. AMPLIFIED VOICE COMMUNICATION, EARPHONE LISTENING

What about systems utilizing electronic amplification and, in particular, radio systems where distance between talker and listener is of no import? Several flight deck radio systems designed for use in high-level noise have been evaluated, using the same talker and listeners as used in the face-to-face and sound-powered-equipment tests.[3,4,7] Reference back to the right-hand side of Fig. 1 shows the limiting results of such talking-listening tests. Although the details are not shown in Fig. 1, in general it can be said that, when a microphone is in relative quiet (or is well-shielded) or when the listener is in quiet, i.e., in Q-N or N-Q conditions, satisfactory military-word intelligibility (greater than 70% rhyme word score) prevails up to levels greater than 125-dB(C) jet noise. Only when both the talker and the listener are in noise (N-N, the data shown in Fig. 1) is the intelligibility limited by 125-dB(C) jet noise, and in this case optimum intelligibility results from listening via earphones in muffs as good as those with liquid-filled cushions.[23]

What techniques, precautions, and circuitry can be utilized to optimize talking and listening in noise? The major objective is to maintain a sizeable differential between the speech signal and the unwanted noise at the ear of the listener.

To maintain a satisfactory speech-to-noise differential at the input of a communication system operated in noise, it is necessary that the talker increase his vocal output as the ambient noise around him increases. Tests in noise, with talkers using noise-shielded microphones and wearing muffs over earphones, indicated that the level at which the talker hears his own voice has considerable effect on his vocal output.[3] Results for a single talker showed that maximum vocal output was obtained when no sidetone was provided. As sidetone level was increased, the talker reduced his vocal effort. Tests with several talkers showed that the intelligibility of speech from talkers in noise did not change significantly as sidetone level was varied, except when sidetone was 10 dB over the preferred level. With more than the preferred amount of sidetone, talkers reduced their vocal effort, the S-N dropped, and intelligibility was adversely affected. Apparently, for less than preferred amounts of sidetone, gains in speech-to-noise differentials counterbalance any distortion incurred as the talker approaches maximum vocal effort, even though Pickett[24] shows that, other things being equal, intelligibility decreases as maximum vocal effort is reached.

In radio systems, if the speech is well above the noise in the received and rectified radio-frequency signal, the speech should be intelligible if the ambient noise leaking through the earmuffs does not mask out the speech from the earphones. It would appear that, if too much of this noise leaks to the listener's ears, a satisfactory S/N could be preserved by merely increasing the amplitude of the speech signal from the earphones. Several factors limit the practicality of this solution. The principal factor is comfort: people do not like extremely loud sounds, whether they be speech or noise. A second factor is safety: loud sounds can and do cause hearing losses if listened to over prolonged periods.

Concerning the comfort aspect of listening to loud speech sounds, two processing techniques can be used:

(1) A noise-actuated automatic volume control (AVC) can be used so that the voice level in the earphones will always be at a comfortable and intelligible level relative to the ambient-noise background. Pollack[25] has shown that such an AVC does not interfere with, and may substantially improve, speech intelligibility. A small microphone in the noise field to which the listener is exposed can provide the noise signal to activate the AVC circuit. Use of the AVC circuit has the additional advantage of simplifying the system from

[23] For noise-attenuation results on many of the latest ear-protective muffs, see Ref. 11. When used in sound-powered equipment, see Ref. 32 and P. O. Thompson, "Acoustic Attenuation of Certain Hearing Protectors and Sound-Powered Telephone Sets," NEL Tech. Mem. 710 (17 July 1964).

[24] J. M. Pickett, J. Acoust. Soc. Am. 28, 902-905 (1956).

[25] I. Pollack, J. Acoust. Soc. Am. 29, 1324-1327 (1957).

the user's standpoint by eliminating a manual volume control. The design of an AVC circuit can be based on the data shown in Fig. 3. For the curves shown in Fig. 3, a group of listeners wearing earmuffs over earphones adjusted the speech level from their earphones to three criteria: maximum level that they would tolerate, level preferred, and minimum level for 80%–90% rhyme-word intelligibility.

(2) Peak clipping can be used at the higher listening levels. For example, up to 12 dB of clipping has a negligible effect on speech intelligibility, does not destroy voice quality, and protects the listener's ears from uncomfortable and potentially deafening speech peaks.[36]

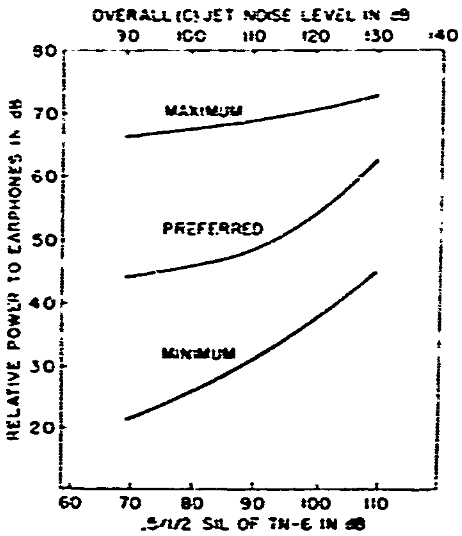


FIG. 3. Relative earphone listening level vs noise level for acceptable listening levels.

Both of these techniques have been used in the flight-deck radio developed at the U. S. Navy electronics Laboratory.[1,37]

Concerning the second, or safety, factor relative to amplifying speech indefinitely to make it intelligible in high-level noise: How loud can a sound be before hearing losses are a potential problem? Flanagan and Guttman[38] have proposed that (for exposures over the years during work-week hours), if the C and A levels are within 1 dB of each other, a C (or A) level greater than 80 dB should be regarded as unsafe. If the C–A level difference is between 1 and 6 dB, a level of 85 dB(C) should be considered unsafe. If C–A is between 6 and 12, 90 dB(C) is the limit, and if C–A is greater than 12, 95 dB(C) is the safe–unsafe boundary. Young[39] has suggested that this is almost tantamount to saying that if the A level is less than 80 dB the noise is safe.

If noise levels greater than 80 dB(A) for a working day are potentially unsafe,[40] then speech levels (at the ear) that continually exceed ambient noises at these levels would also be unsafe.

In general, good speech intelligibility in high-level noise has been obtained[7,37] by using (1) a noise-canceling microphone (in a noise shield[41,42] for extreme noise); (2) earphones in noise-attenuating earmuffs,[11] (3) a wide speech bandwidth (three or more octaves wide), centered somewhere between 1000 and 1800 cps (Ref. 43); (4) minimum or no sidetone[1]; (5) AVC circuit to conform in general to preferred listening level as shown in Fig. 3; (6) peak clipping of 12 dB at maximum power[1,36,37]; and (7) flat frequency response and minimum distortion in the audio circuitry.[37]

It is often thought that speech picked up at some anatomical location other than in front of the lips would aid speech intelligibility in noise. Studies[44–47] concerned with vibratory or bone pickup show in general that this is not so and that, as the distance from the larynx increases, the amplitude of picked-up sound decreases. Likewise, ear-insert microphones work well in the quiet[48] but are not in competition with microphones located in front of the lips in high levels of noise.[49]

An evaluation of representative vibratory and ear-pickup microphones in 1957 stated[47] that, "Transducer outputs are intelligible . . . to the following sound-pressure levels [of jet noise]: 137 dB for the M-55 [noise-canceling dynamic microphone in noise shield]; 120 dB for the D-98 Airphone [an ear-insert phone]; 110 dB for the PDR-8 [worn over the ear]; 120 dB for the VM-1 [vibration pickup] on the forehead; and 130 dB for the VM-1 on the mandible." Subtract 21 dB from all these values to equate to a 0.5/1/2 SIL of equivalent flat (or minus 6 dB/oct) thermal noise.

A comprehensive U. S. Air Force-sponsored study by the Western Electro-Acoustic Laboratory on the per-

[7] J. C. R. Licklider and G. A. Miller, "The Perception of Speech," in *Handbook of Experimental Psychology*, S. S. Stevens, Ed. (John Wiley & Sons, Inc., New York, 1951), Chap. 26, pp. 1041–1062.

[37] D. C. Gibson, "USNEL Flight Deck Communications System. Part I. Noise Attenuating Radio Communications Helmet and Fixed Station Equipment," NEL Rept. 922 (14 Mar. 1960), AD 254 531.

[38] J. L. Flanagan and N. Guttman, J. Acoust. Soc. Am. 36, 1654–1658 (1964).

[39] R. W. Young, personal communication and footnote on Table III, Ref. 38.

[40] The Revision of Document 43 (Secretariat 194) 314 of Technical Committee 43, Working Group 8 of the International Organization for Standardization, dated June 1964, puts this safe-unsafe boundary level at 90 dB(A) for 5-h days.

[41] M. S. Hawley, J. Acoust. Soc. Am. 28, 1256–1260 (1956).

[42] M. S. Hawley, J. Acoust. Soc. Am. 30, 188–190 (1958).

[43] J. C. Webster, "Generalized Speech Interference Noise Contours," J. Speech Hearing Res. 7, 133–141 (1964).

[44] J. Muliendore, "An Experimental Study of the Vibration of the Bones of the Head and Chest during Sustained Vowel Sounds," Speech Monographs 16, 163–176 (1949).

[45] G. von Békésy and W. A. Rosenblith, "The Mechanical Properties of the Ear," in *Handbook of Experimental Psychology*, S. S. Stevens, Ed. (John Wiley & Sons, Inc., New York, 1951), Chap. 27, p. 1111.

[46] H. Moser and H. J. Oyer, J. Acoust. Soc Am. 30, 275–277 (1958).

[47] J. C. Webster, P. O. Thompson, J. Snidecor, and D. D. Washburn, "Speech Pickup from Various Anatomical Locations," Proc. Decade Basic Appl. Sci. in the Navy, Off. Naval Res., Washington, D. C. (1957).

[48] H. J. Oyer, J. Acoust. Soc. Am. 27, 1207–1212 (1955).

[49] R. D. Black, J. Acoust. Soc. Am. 29, 260–264 (1957).

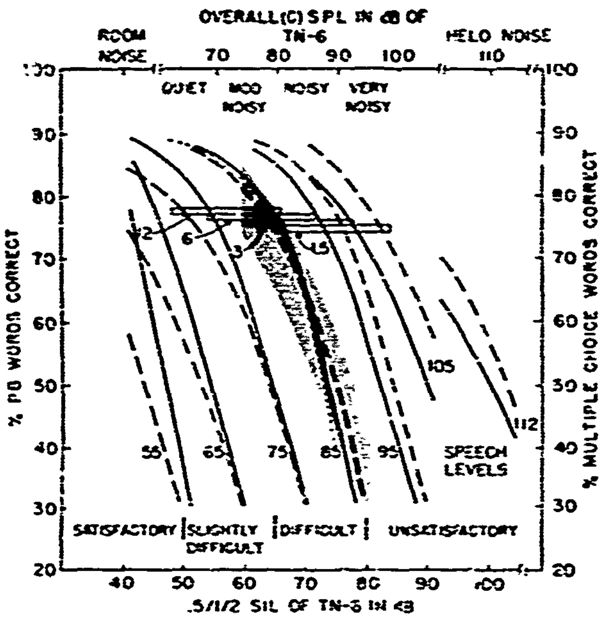


FIG. 4. Speech intelligibility (percent PB words correct) for speech at various levels vs noise level of a −6-dB/oct noise (Fig. 2, Ref. 23) and (at extreme right) a hovering HUP helicopter noise (relative octave levels of −23, −8, −9, −7, 0, −3, −8, and −16 dB for octaves centered at 67, 125, etc.). The solid lines are for open ears, the dotted lines for protected ears (plugs or muffs). The majority of the data in this Figure are replotted from Fig. 3, Ref. 23 (with abscissa and parameter values interchanged). The data at the extreme right are from Ref. 6. To equate the two sets of data, use is made (on the bottom abscissa) of the 0.5/1/2 SIL of −6-dB/oct thermal noise. No equating is done on the ordinate re the comparative intelligibility of PB words[23] (left) and multiple-choice words[22] (right).

formance of electroacoustic transducers in noise has been reported orally[50–52] but as yet appears only in company reports,[53] and the data are too extensive to summarize here. It was concluded that a microphone embedded in a dental plate or otherwise fastened to the lower teeth inside the mouth showed promise for communicating in high noise levels.[53]

For communicating in high levels of noise, the old reliable carbon button has had its day. At ambient levels above 80 dB over-all jet noise (50-dB equivalent 0.5/1/2 SIL of flat TN and TN-6), dynamic (and condenser) microphones will give better intelligibility scores.[6,8,9]

V. AMPLIFIED VOICE COMMUNICATION, LOUDSPEAKER LISTENING

What are the limiting levels of noise in which amplified speech can be heard over a loudspeaker? Figure 4 is a replot of some data of Kryter's (Fig. 3 in Ref. 23). In this replot, the abscissa and parameter dimensions are interchanged from Kryter's original Fig. 3. In this form, it is easy to see how speech intelligibility for loudspeaker-reproduced amplified speech is limited by ambient noise and how this varies with (1) amplified speech level and (2) with the wearing of earplugs (dotted lines), versus open ears (solid lines). The remaining alteration to the original Kryter data is that noise level is now plotted as the 0.5/1/2 SIL (of the −6-dB/oct thermal noise) instead of as the over-all level.

When these two transformations are made, many comparisons are possible. For example, note the shaded areas around the 85-dB speech-level curves. This area delimits Kryter's data on his Fig. 7 where talkers and listeners were in a face-to-face communication situation at a distance of 7 ft. The talkers (just as the talker used to generate the data in Fig. 1) "spoke at whatever voice level they thought necessary to make themselves understood" (Ref. 23, p. 416). The upper edge of the shaded area is when neither talkers nor listeners wore earplugs and the lower boundary is when both talkers and listeners did wear earplugs. The original data show points in between, where one or the other wore plugs.

The 6-ft curve from Fig. 1 (lowered 10 percentage points to account for the difference between PB words correct and rhyme words correct,[2] i.e., 50% PB score = 60% rhyme score) would fit in this same shaded area. Beranek's[17] 8-ft, 86-dB speech level, treated-aircraft-cabin noise curve would also fall in the shaded area (between the limits of 40% and 80%).

It should be noted in passing that all data in Fig. 4 (and the nonearphone data in Fig. 1) were taken in semireverberant (certainly nonanechoic) rooms. Kryter shows (Ref. 23, Fig. 5) that for open ears the intelligibility in an anechoic room is 10% greater than in his reverberant room. For plugged ears, the 10% holds only at high levels of intelligibility. And Thompson, Webster, and Gales[54] have shown that any amount of reverberation added to "dead" speech causes a decrement in intelligibility.

Also shown in Fig. 4 is a summary of the data relating SIL, voice level, and distance between talker and listener (corrected from 0.6/4.8 SIL to 0.5/1/2 SIL). The same rationale and explanations as used when the SIL data were superimposed in Fig. 1 also apply here.

At the top of the Figure are cutoff points delimiting the subjective opinions of people relating to the relative noisiness of "stenographic and large engineering drafting rooms."[22] [Executive-office personnel tolerate 10 dB less noise; that is, they think that offices are "noisy" at an 0.5/1/2 SIL of 60 dB (not 70 dB), and "very noisy" at an 0.5/1/2 SIL of 70 dB.]

At the bottom of Fig. 4 are the opinions of the people surveyed by Beranek[21,22] relative to measured noise level (here replotted as 0.5/1/2 SIL) and their ability to use the telephone.

[50] W. B. Snow, J. Acoust. Soc. Am. 33, 1661(A) (1961).
[51] P. S. Veneklasen, J. Acoust. Soc. Am. 33, 1661(A) (1961).
[52] J. P. Christoff, J. Acoust. Soc. Am. 33, 1662(A) (1961).
[53] "Study and Investigation of Specialized Electro Acoustic Transducers for Voice Communication in Aircraft," Western Electro-Acoust. Lab., Inc., Final Rept. with Suppl. App. 1-6, Contract AF33(616)-5710, Task No. 43060 (Feb. 1959).

[54] P. O. Thompson, J. C. Webster, and R. S. Gales, J. Acoust. Soc. Am. 33, 604–605 (1961).

To augment the data of Kryter[23] and Beranek,[5,2] some data of Webster, Thompson, and Beitscher[6] are added to the right in Fig. 4. The noise was "hovering helicopter," the words were "multiple choice,"[33] and muffs were used instead of plugs, but the data appear to be consistent with Kryter's[23] and do extend the reproduced level of speech and noise slightly beyond the limits tested by Kryter.

Some general conclusions to be drawn from Fig. 4 and the data on which it is based are. At noise and/or speech levels above some point, intelligibility is better when listeners and/or talkers are wearing earplugs[23] or earmuffs.[6] This limited noise level is roughly at a 0.5/1/2 SIL of 70 dB and the limiting speech level is 95 dB. Since people dislike wearing hearing protectors in marginal levels of noise (and don't believe that they can hear better), a 0.5/1/2 SIL of 80 dB should probably be chosen as a point where some sort of pressure should be brought to bear to insist that ear protection be worn. A 0.5/1/2 SIL of 80 dB corresponds roughly to an A level of 90 dB (see Table I or Ref. 12), where hearing protection should be worn as a protection against noise-induced permanent hearing loss with 8-h, 5-day, 50-week, multiple-year exposures.[35-40]

Another conclusion from Fig. 4 and its sources[6,23,2] is that at any constant speech-to-noise differential the intelligibility is remarkably constant, rising some at low noise levels and falling some at high noise levels (even when wearing ear protectors). This slight falloff of intelligibility with level has been found by others[34] and suggests that the equivalent to an ear-overload condition exists, just as Pickett[26] has shown a similar voice overload (or at least a decrement of intelligibility with greatly increased vocal effort).

It is interesting to note that "noisy" and "very noisy" spaces are associated with SIL's such that "shouting" or "very loud" voice levels (or 95-dB speech levels) are required for conversations at 1.5 or 3 ft, and this is the region where telephone conversations are judged to be "difficult" or "unsatisfactory." All of these adverse noise conditions occur at the region where ear protection will aid intelligibility and at the boundary where ear protection should be used to protect against hearing losses. It is quite apparent that 0.5/1/2 SIL's above 70 dB should be avoided in spaces where people must converse or communicate via some communication device. Spaces where conversations cannot be carried on in comfort at 3 ft are too noisy for any type of task requiring face to face communications. Noiseproofing of humans and equipment should start at this noise level. At 0.5/1/2 SIL's greater than 90 dB the wearing of hearing protection should be mandatory and every noiseproofing technique (except a noise shield for the microphone) should be employed. At 0.5/1/2 SIL's above 100 dB, every noiseproofing technique should be employed.

Beranek's[2] recommendation that 0.6–4.8 SIL's above 55 dB (0.5/1/2 SIL's of 60 dB) should be avoided for office spaces appears to be a good choice. Levels 10 dB greater than this should be avoided even on ships where ventilation and other noises are ever present and their quieting expensive.

ACKNOWLEDGMENTS

Although this paper has been an attempt to draw together some aspects of speech intelligibility in noise, much reference material was not used. My apologies are extended to K. D. Kryter, Irwin Pollack, L. L. Beranek, J. M. Pickett, P. S. Veneklasen, J. W. Black, and others too numerous to mention, for not adequately covering and/or acknowledging their past work in this field. I do wish to acknowledge the contributions of P. O. Thompson and R. G. Klumpp in particular, and others at the U. S. Navy Electronics Laboratory in general, for the many hours of work expended in evaluating equipments and serving as experimenters (and subjects) to generate the data that have been summarized here.

---

[33] J. W. Black, "Multiple-Choice Intelligibility Tests," J. Speech Hearing Disorders 22, 213–236 (1957).
[34] I. Pollack and J. M. Pickett, J. Acoust. Soc. Am. 30, 127–130 (1958).

REFERENCES

1. Klumpp, R. G. and Webster, J. C., "Predicting Speech Interference From Physical and Psychophysical Measures of Ambient Noise," in Speech Communications Seminar, Stockholm, 1962. Proceedings, September 1962

2. Bureau of Ships, General Specifications for Ships of the United States Navy, Section S1-10, "Noise, Shock, and Vibration," 1 May 1958

3. Military Specification MIL-E-22843 (Ships), Equipment, Low Noise Level (Naval Shipboard Use), 15 March 1961

4. Webster, J. C. and Klumpp, R. G., "Effects of Ambient Noise and Nearby Talkers on a Face-to-Face Communication Task," Acoustical Society of America. Journal, v. 34, p. 936-941, July 1962

5. Thompson, P. O. and Webster, J. C., "The Effect of Talker-Listener Angle on Word Intelligibility," Acustica, v. 13, p. 319-323, 1963

6. Thompson, P. O. and Webster, J. C., "The Effect of Talker-Listener Angle on Word Intelligibility, II: In an Open Field," Acustica, v. 14, p. 44-49, 1964

7. American Standards Association Standard S1.4 - 1961, General-Purpose Sound Level Meters, Specification For, 1961

8. Octave-Band Sound-Pressure-Level (SPL) Measurements Made on the USS CORAL SEA (CVA 43), 6 September 1961 Through 13 September 1961, Enclosure (1) to U. S. Naval Reserve Research Company 12-5 Letter Report to COMNAVAIRPAC dated 4 January 1962 Forwarded by COMNAVAIRPAC Letter FF4-1/9607/2, Serial 7314/771 to Navy Electronics Laboratory, dated 25 January 1962

9. Bureau of Ships. Code 371 Report 371-N-12, Criteria for Setting Airborne Noise Level Limits in Shipboard Spaces, by M. Strasberg, 15 July 1962

10. New York Naval Shipyard, Brooklyn. Material Laboratory Report 5280-1; Final Report, Report of Investigation of Airborne Noise in C.I.C., Sonar, and Chart Room Spaces, USS BORIE (DD 704), by S. Feldman, 5 March 1952

10. Continued

New York Naval Shipyard, Brooklyn. Material Laboratory Report 5280-2; Final Report, Report of Investigation of Compartment Airborne Noise Survey Techniques, USS MASSEY (DD 778), by K. R. Mercy and M. Bernstein, 28 May 1953

New York Naval Shipyard, Brooklyn. Material Laboratory Report 5280-6; Final Report, Report of Airborne Noise Survey in USS TIMMERMAN (EAG 152), by A. Simowitz, 26 September 1956

New York Naval Shipyard, Brooklyn. Material Laboratory Report 5280-19; Final Report, Report of Survey of Airborne Noise on USS FORRESTAL (CVA 59), by K. R. Mercy, 2 August 1956

New York Naval Shipyard, Brooklyn. Material Laboratory Report 5280-23; Final Report, Report of Investigation of Airborne Noise on U. S. Army Ferry LT. SAMUEL S. COURSEN (FB812), by A. N. Savacchio and M. M. Kushnick, 1 February 1957

11. Southwest Research Institute Contract NObs 88636; Final Technical Report, Study of Airborne Noise From Shipboard Machinery, by C. R. Sparks and others, 19 June 1964

12. Netherlands' Research Centre T.N.O. for Shipbuilding and Navigation Report 44S, Some Acoustical Properties of Ships With Respect to Noise Control, Part 1, by J. H. Janssen, August 1962

Netherlands' Research Centre T.N.O. for Shipbuilding and Navigation Report 45S, Some Acoustical Properties of Ships With Respect to Noise Control, Part 2, by J. H. Janssen, August 1962

13. Skipsteknisk Forskningsinstitut Report 35, The Acoustical Condition on Board Modern Merchant Ships (English Title), by J. Storm, January 1962

14. Skipsteknisk Forskningsinstitut Report M47, The Noise Conditions on Board Merchant Ships Delivered in 1962 (English Title), by J. Storm, March 1963

15. Kallenbach, W. and others, "Untersuchungen zur Hörbarkeit von Schallsignalen auf Seeschiffen," Acustica, v. 11, p. 283-290, 1961

16. Hawkins, J. E. and Stevens, S. S., "The Masking of Pure Tones and of Sp ch by White Nois," Acoustical Society of America. Journal, v. 22, p. 6-13, January 1950

17. Greenwood, D. D., "Auditory Masking and the Critical Band," Acoustical Society of America. Journal, v. 33, p. 484-502, April 1961

REVERSE SIDE BLANK

Reprinted from The Journal of the Acoustical Society of America, Vol. 34, No. 7, 936-941, July, 1962

Effects of Ambient Noise and Nearby Talkers on a Face-to-Face Communication Task

John C. Webster and Roy G. Klumpp
U. S. Navy Electronics Laboratory, San Diego 52, California
(Received February 26, 1962)

From 1 to 5 talker-listener pairs, talkers seated shoulder-to-shoulder on one side of a table with listeners on the other, communicated word lists in conditions of quiet and ambient thermal noise levels of 65, 75, and 85 dB. Each talker read one word at a time to his listener-partner, who repeated back each word for verification by the talker. Talker-listener pairs were instructed to maintain an accuracy of 90% or better. For the lower ambient levels the speech level of a central pair increased about 5 dB for an additional 10 dB of noise or for each doubling of the number of pairs around them. The rate of utterance decreased with noise but showed no clear-cut pattern of change as the number of additional talkers was varied. Accuracy of communication was, on the average, 94% and was never below 84%. Communication errors defy simple description but in general (1) for a constant noise level, increasing the number of talkers results in increasing errors; and (2) for 3 or fewer talker-listener pairs, percent error does not increase until the ambient-noise level reaches 85 dB.

INTRODUCTION

It is often necessary to assess the suitability of a room (existing or proposed) for speech communication, from a knowledge of measured or expected ambient-noise levels in the room. The prediction is as often as not based solely on physically existing noises (fans, vehicular noise, etc.) and may not adequately account for the cumulative noise effects due to the people within the room performing voice communication tasks. The purpose of this study is to determine the interaction between physically introduced ambient noise and speech levels of nearby talkers on the performance of talker-listener pairs performing a communicating task (exchanging Harvard PB words[1]). The variables measured are: speech level required to communicate in increasing levels of physical and/or talker-generated noise, rate of utterance, and the number of errors.

APPARATUS

All testing was done in a 28×16×10 ft recording studio. "Quiet" levels in the studio were 55 dB or less on the C scale (35 dB or less on the A scale) of a sound-level meter. Reverberation time of the unoccupied studio was approximately 0.25 sec.

The talker and the listener of a central pair were seated across from one another at the center of a table measuring 60×34 in. When a second and/or third pair was added, they were placed beside the center pair so that all talkers were on one side of the table and all listeners on the other. Talkers were seated shoulder-to-shoulder and listeners were shoulder-to-shoulder. Mouth-to-ear distance between any talker-listener pair was 36 in. Center-of-head-to-center-of-head distance between adjacent talkers (or listeners) was approximately 17 in. Pairs 4 and 5, when used to generate additional babble, sat at two small additional tables placed one at each end of the larger central table.

Ambient noise was obtained from three speaker arrays driven by amplifiers and a thermal noise generator. The noise field about the table was uniform to within ±1 dB as measured on the C scale of a sound-level meter. The frequency response of the ambient-noise system was ±5 dB from 300 to 6000 cps as measured with warble tones and a microphone located above the center of the table.

Six matched microphones were used to record the outputs of the six subjects (3 talkers and 3 listeners) from whom data were taken. Microphones were fastened to the table in front of these six subjects. Each person positioned himself so as to have his lips almost touching the edge of the microphone. Lip distance to center of microphone was approximately 1 in. Microphones were oriented so as to minimize breath blast. Outputs of the six microphones were recorded on tape and were later analyzed for error, rate of utterance, and speech level.

Speech level was determined by playing the tapes back at half the recorded speed and tracing the speech signals on a Bruel and Kjaer, type 2304 power level recorder using a writing rate of 1000 dB/sec. Sound-pressure levels were assigned to these speech tracings by determining the difference between the speech peaks and the background-noise level. The background-noise level had previously been measured with a sound-level meter set to C, "fast."

The level of each word was noted to the nearest decibel and the median level for a list was determined. Medians were used instead of means because of the reduced analysis time required. Later analysis showed that medians differed from mean values, in 24 lists sampled, by less than ±1 dB. In general the quartile values differed from the median by about 1.5 dB and never differed by over 5 dB.

PROCEDURE

From 2 to 10 people were used as subjects, always acting in pairs. They were members of the U. S. Navy Electronics Laboratory, had normal or near normal

[1] J. P. Egan, Laryngoscope 58, 955-991 (1948).

hearing, and were experienced in noise communication tests. One member of the pair, the talker, read words, one at a time, from a list of PB words. His partner, the listener, repeated back each word as he heard it. If the talker was satisfied that the listener had repeated the word correctly, he proceeded on to the next word. If not satisfied as to the correctness of the response, the talker, acting on previous instructions, repeated the word and again ascertained the listener's response. Each pair was instructed to maintain a high degree of accuracy (90% to 100%) but not to repeat any given word more than three times. Maintaining the required accuracy was the responsibility of the talker, who had 15 words before each test list to adjust his vocal output until he was satisfied with the level of accuracy of his listener's responses. He was cautioned by the experimenter, who monitored each list, that any list with excessive errors would be rerun. No lists were actually rerun (although scores between 84 and 90% did occur on 2½% of the lists). They were also instructed to deliver the words as rapidly as possible while maintaining required accuracy. This was done to minimize the possibility of getting in synchrony and delivering words in the interval between words of an adjacent pair. Subjects were instructed to ignore, as best they could, the ambient noise and the speech from adjacent pairs. No instructions were given as to exactly what vocal output would be required.

The results of 3 pairs were scored although 5 pairs were used in some parts of the experiment. Pairs 4 and 5 performed the same tasks as the three centrally located pairs but their results were not scored. They served simply as "noise" sources.

There were four groupings of talker–listener pairs: each pair alone, two pairs together, all three data-yielding pairs, and the three data-yielding pairs plus two more babble-contributing pairs. There were 10 experimental sessions; each pair alone (3) replicated once (total of 6), pair 1 vs pair 2, pair 1 vs pair 3, pairs 1, 2, and 3 together, and pairs 1, 2, 3 plus pairs 4 and 5. Within each of the 10 sessions, tests were carried out in the following order of noises: Quiet, 65, 75, 85, 85, 75, 65, Quiet.[2] Each test consisted of reading 75 PB monosyllabic words (but scoring only the middle 50 words which comprised a complete PB list). The words preceding the actual test words allowed a "settling in" period for adopting a mutually acceptable level and rate. The words following ensured a constant babble level for the pairs that were still communicating after the faster pairs had finished.

The experiment was designed so that (1) no adjacent pair read the same list at any time during a session, (2) the extra words at the beginning and end of a list were taken from lists not used in that particular test session, and (3) during any given test session each pair read a different list for each condition. Preliminary runs indicated that seeing the talker's lips made no difference in any criterion measure of accuracy, rate, or speech level, and so no screen was used between talkers and listeners.

[2] In the condition where 5 pairs were communicating, only 4 noise conditions were used (the ascending sequence of Quiet, 65, 75, and 85 dB).

	SEQUENCES								ERRORS			
	T	L	T	L	T	L	T	L	LE +	TOE +	TCE =	ΣE
1	A	A							0	0	0	0
2	A	B	A	A					1	0	0	1
3	A	B	A	C	A	A			2	0	0	2
4	A	B							1	1	0	2
5	A	A	A	A					0	0	1	1
6	A	B	A	A	A	A			1	0	1	2
7	A	A	A	B	A	A			1	0	1	2
8	A	A	A	B	A	C	A	A	2	0	1	3

A = CORRECT WORD, B & C = INCORRECT WORDS, T = TALKER, L = LISTENER
LE = LISTENER ERROR, TOE = TALKER OMISSION ERROR, TCE = TALKER COMMISSION ERROR
ΣE = TOTAL ERRORS

FIG. 1. Types of errors made by talkers T and listeners L. The eight word sequences shown at the left encompassed every type of error found in these tests. The type of error and the way these errors were summed are shown at the right.

TREATMENT OF DATA

The raw data consisted of magnetic tape recordings of three talkers and three listeners. These lists of 50 words were analyzed for (1) speech level, (2) rate, based on the time taken for the talker to say and the listener to repeat 10 consecutive words in a stable (error-free, repeat-free) portion of their performance, and (3) number of errors.

To understand the variety of possible errors, refer to Fig. 1. For any one word the sequence of events which might transpire between talker T and listener L is shown in Fig. 1. The symbol A represents the correct word while B and C represent incorrect responses to A. Actually the talker by definition never said an incorrect word. If he misread a word on his list, his misreading became the new correct word. On the average, 94% of the responses were correct, that is, sequence 1. The most common incorrect sequence was sequence 2, which sometimes got lengthened into sequence 3. Sequences 2 and 3 involve listener errors LE but no talker errors. Sequence 4, which occurred most typically in the higher ambient-noise conditions, involves an error on the part of both the listener and the talker. The talker's error is that he failed to detect the listener error and therefore failed to correct it (TOE or Talker Omission Error). Sequence 5 is purely a talker error: He thought the listener's correct response was incorrect and repeated a word he need not have repeated (TCE or Talker Commission Error). This type of error sometimes sets off a longer chain of errors like sequences 7 and 8 where the listener, because of the error of the talker, assumed he was wrong and changed his initially correct response to one (7) or more (8)

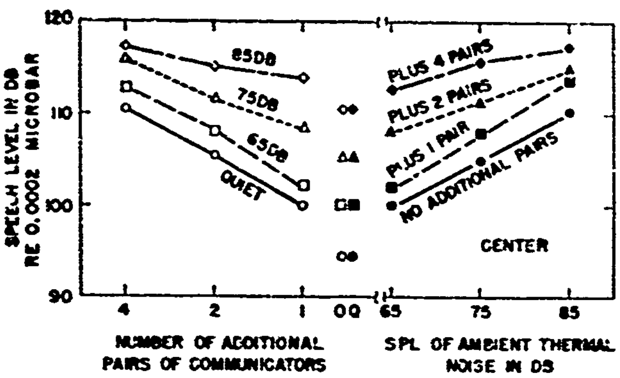


FIG. 2. Speech level vs noise level of the center pair only. The midpoint along the abscissa represents the quietest condition, a single talker–listener pair in the quiet recording studio. To the right of center, thermal noise was added to make ambient room noise levels at 65, 75, and 85 dB (C scale). To the left the number of communicating pairs was successively doubled by the addition of 1, 2, and 4 pairs of talker-listeners. The parameters on the left hand side of the figure represent greater noise levels, and on the right, more people. The same data are plotted on the left and right halves. The two halves differ only in that the abscissa and parameter designations are interchanged.

The datum points for "no additional pairs" and "plus 1 pair" represent results from 8 different lists of 50 PB words. Four lists are represented for "plus 2 pairs" and two lists for "plus 4 pairs."

incorrect responses. Sequence 6 is another, though rare, combination of listener and talker errors. In plotting results, all three types of errors were combined into a single error score, i.e., E=LE+TOE+TCE.

The number of repeats is ordinarily an important measure of communications success. However, since the lack of a repeat, sequence 4, is as serious as the presence of a repeat, the total number of repeats is not a good performance measure on this task. The information on repeats, or lack of them, is contained in the combined error score.

RESULTS

Speech Level

When averaged over all three pairs, the speech levels of the listeners did not differ from the speech levels of the talkers. Therefore in the results the speech levels of the talkers and the listeners are averaged.

Figure 2 shows how the speech level of the center talker–listener pair changed with ambient thermal noise level and number of adjacent competing talker–listener pairs. Median speech level is plotted on the ordinate and along the abscissa are measures, direct or indirect, of noise levels. The speech levels on the ordinate are levels measured approximately 1 in. from the talker's mouth (at the listener's ear, 36 in. away, the measured level was 18 dB less). The minimum speech level of 94 dB (center pair alone in the quiet) corresponds to "normal" speaking effort and the maximum of 117 dB ("plus 4 pairs" and "85 dB") to a "shout." The true long-term rms speech level at 1 m in a free field would be 34 dB less than the 94 dB (normal) and 117 dB (shout) values shown (31 dB for distance doubled and 3 dB for peak to rms correction of the sound-level meter reading).

From the center to the right along the abscissa the ambient room noise increases in 10-dB steps. The break in the abscissa reflects the fact that, although the C scale level for the "quiet" Q condition is 10 dB below the 65-dB thermal-noise level, the difference in A-scale levels is about 30 dB. The A level is more closely related to speech interference than is the C level.[3] The parameters on the curves to the right are the number of additional pairs communicating: The curves are (1) the center pair by themselves, (2) with 1 additional pair alongside, (3) with 2 additional pairs (one on each side of the center pair), and (4) with 4 additional pairs (two pairs on each side of the center pair).

From the center to the left the number of pairs of communicators around the center pair doubles and the parameters on the curves are the ambient-noise levels. The same data are plotted on the left as on the right. The difference is that the abscissas and parameters are interchanged.

This symmetrical plotting of the data is done to emphasize the comparison between the effects on speech level of "noise level" and "number of communicators." The curves are within ±2 dB of being bilaterally symmetrical around the center. This implies a general equivalence between (1) adding 10 dB of noise and (2) doubling the number of people. Adding 10 dB of noise essentially doubles the subjective loudness.[4] If speech levels of talkers surrounded by noises be considered an indicator of "loudness" then doubling the number of communicators likewise doubles the loudness. In the lower ambient-noise levels the amount of speech-level shift with 10 dB of noise, or doubling of communicators, is roughly 5 dB. In the highest ambient-noise conditions, speech level is approaching maximum possible sus-

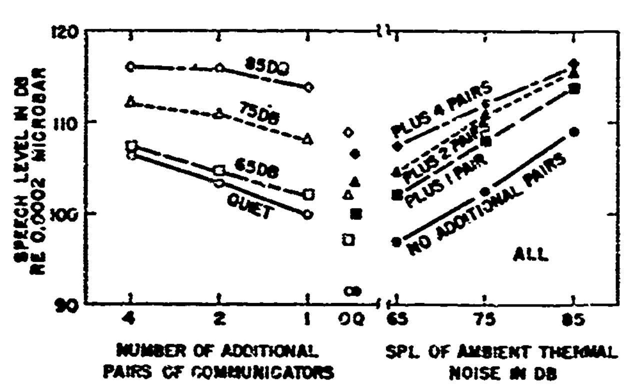


FIG. 3. Speech level vs noise level of the 3 data-yielding pairs (all). The ordinate, abscissa, and parameters have the same meaning as in Fig. 2. Each datum point on the "no additional pairs" represents 24 lists, on the "plus 1 pair" 16 lists, on the "plus 2 pairs" 12 lists, and on the "plus 4 pairs" 6 lists.

[3] R. G. Klumpp and J. C. Webster, "Speech Interference Aspects of Navy Noises," U. S. Navy Electronics Laboratory Tech. Rept. (to be published).

[4] S. S. Stevens, J. Acoust. Soc. Am. 28, 807–832 (1956). See Fig. 19.

tained level[5,6]: Physiological limitations keep the speech level from continuing to increase linearly.

An increase in ambient thermal noise affects all talkers in the room equally, but adding people symmetrically around the center pair affects the center pair more (or at least differently) than any of the other pairs. In Fig. 3 the speech levels of all 3 pairs are considered. The same general trends are evident in Fig. 3 as in Fig. 2; i.e., speech level increases with increasing ambient noise or additional adjacent talkers. Two kinds of differences show up. First, the inherent speech levels of the two noncentral pairs are less than the center pair, so with "no additional pairs" the speech level is down about 4 dB on Fig. 3.

The second difference has to do with geometry of the test situation. When alone or with one additional pair, the center pair and the two noncentral pairs are in essentially the same test situation. However, with two additional pairs the center pair has a pair on each side while each noncentral pair has the additional people on one side of them only. To the noncentral pairs, the "plus 2 pairs" condition is more like "plus one and a fraction." The results reflect this inasmuch as the "plus 2 pairs" data on Fig. 3 lie between the plus 1 and the plus 2 of Fig. 2. Likewise with four additional pairs in the room the center pair is surrounded by two pairs on each side whereas the noncentral (but nonend) pairs have three pairs on one side of them and one pair on the other. The "plus 2 pairs" condition for the center pair is more equivalent to the "plus 4 pairs" for the noncentral pairs, and the data reflect this. If the left-hand abscissa of Fig. 3 (and the right-hand parameters) were relabeled as "no additional pairs," "plus 1," "plus 1+," and "plus 2+," the results on Fig. 3 would (except for the inherent weaker speech levels alone) very nearly coincide with the results shown on Fig. 2; the agreement is within 2 dB.

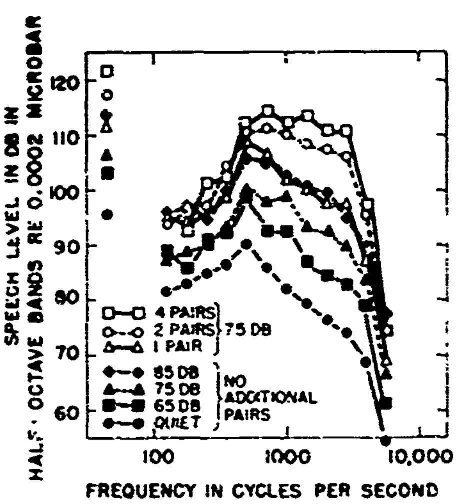


FIG. 4. Voice spectra of the talker from the center pair at various noise conditions. Over-all (broad-band) levels are shown for each condition at the extreme left. These over-alls represent the talker portion of the data plotted on Fig. 2 at the specified parametric points.

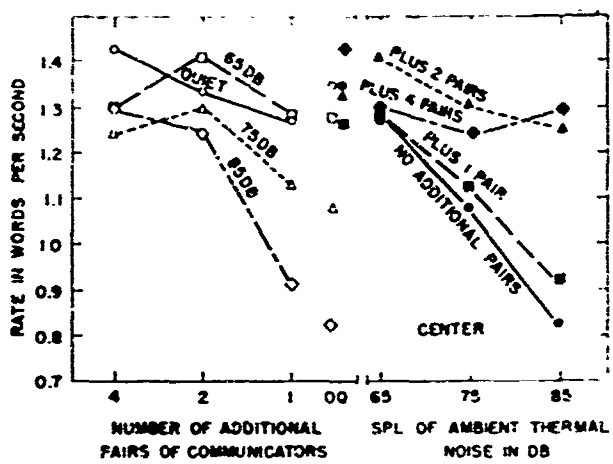


FIG. 5. Rate in words per second to read 10 error-free words as a function of ambient noise and number of communicators. Abscissa and parameters have same meaning as in Figs. 2 and 3, and the data are based on the same number of observations as in Fig. 2.

Figures 2 and 3 show how the over-all speech level increases with noise, but what happens to the voice spectrum as the over-all speech level increases? The voice spectrum of the talker of the center pair was analyzed in detail from selected samples of seven of the different experimental conditions shown in Fig. 2. Figure 4 shows the level in half-octave bands under the conditions of Quiet and at 65, 75, and 85 dB of ambient noise, and in the 75-dB level with 1, 2, and 4 pairs of communicators around him. His over-all level increases in order through each of the conditions just listed (see Fig. 2) except for the "no additional pairs" at 85 dB vs the "plus one" at 75 dB, which are roughly equal. Note that as the over-all level increases more and more energy is shifted to the higher frequencies.[7] The two levels which in over-all level are roughly equal (alone at 85 and 1 pair at 75) have nearly identical spectra. From these selected samples it appears that spectrum is more dependent on level than upon what causes the level to be where it is (i.e., it makes little difference whether it is thermal ambient noise or fellow communicating colleagues that cause increased speech level; for equal speech level the spectrum is the same).

Rate

Figure 5 shows the same type of plot as Fig. 2 except that rate (words per second) is plotted on the ordinate. Note the complete lack of symmetry as con-

[5] J. M. Pickett, J. Acoust. Soc. Am. 28, 902-905 (1956). See Fig. 1.

[6] Western Electro-Acoustics Lab. Final Rept. Contract AF 33(616)-3710 Task No. 43060, "Study and Investigation of Specialized Electro-Acoustic Transducers for Voice Communication in Aircraft." Appendices 1-6, (February 1959). See Fig. A4 9.

[7] J. C. R. Licklider, M. E. Hawley, R. A. Walkling, J. Acoust. Soc. Am. 27, 207(A) (1955).

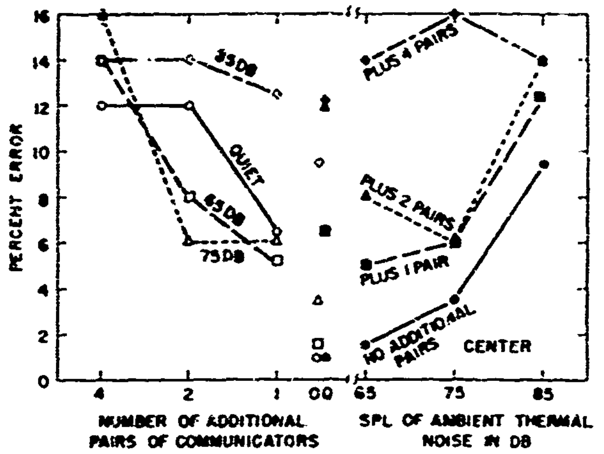


FIG. 6. Errors of the center pair as a function of ambient room noise and number of competing communicators. Abscissa and parameters as in Figs. 2, 3, and 5. Data are based on same number of observations as in Fig. 2.

trasted to Fig. 2. The rate of utterance of the center pair depends to a greater degree on ambient noise than on talker-generated noise. Contrary to the speech-level results, increasing the ambient noise decreases the rate of talking. In general, increasing the number of people increases the rate of talking, but there are strong interactions with room noise. The rate data for "all" communicators are not shown inasmuch as they do not show any systematic deviation from the "center" data.

Errors

Figure 6 shows the percentage of errors made by the center pair. In low noise levels the addition of communicators contributes greatly to the number of errors. Likewise with few communicators in the room additional noise above 75 dB contributes greatly to errors. The increase from 65 to 75 dB of noise does not in general increase the number of errors.

The "all" data on errors (not shown) are generally equivalent to the "center" data except that in general fewer errors were made by the noncentral pairs. It must be remembered that this experiment was designed to test speech level as a function of noise, error rates kept minimal. The talkers were instructed to keep errors to less than 10% and actually achieved a 6% error rate on the average. It is not too surprising therefore that the analyses of errors show no systematic trends.

DISCUSSION

The increase in speech level of 5 dB for each 10-dB increase of ambient thermal noise (slope 0.5) is somewhat higher than the change in speech level with noise found by Kryter,[8] by Hanley and Steer,[9] by Korn,[10] and by Pickett[11] (slope≐0.3). Although factors such as the speech material, room acoustics, distance between talker and listener, spectrum of ambient noise, and absolute ambient-noise level will affect the speech level, differences in slopes obtained are probably due primarily to differences in feedback provided to the talker and degree of accuracy required. In the Kryter and Pickett studies the listener wrote down what the talker said and correctness of response was ascertained later. In the Korn and in the Hanley and Steer studies there was no objective measure of communication effectiveness available to the talker. In contrast, in the present study the talker knew immediately after each utterance how well his speech was being received and the error rate averaged 6%. This difference in accuracy and immediacy of feedback could account for the greater slope found in the present experiment.

In this study the number of error- (and repeat-) free words spoken decreased with increasing ambient noise. This tendency decreased progressively as more communicators entered into the talking task. Hanley and Steer[12] obtained similar results inasmuch as increasing noise in the headphones of talkers caused the speech level to increase, the words spoken per minute to decrease, and the mean syllable duration to increase. On the other hand, Black[13] found that when talkers *were instructed* to read at four vocal-effort levels ranging from soft to loud that speaking rate increased. Black's results do agree, however, that speech level and the fundamental voice frequency increased as talkers spoke louder and louder. Black's talkers were not raising their speech level to combat noise and retain error-free communication but only upon instructions to do so.

When Peters[14] informed his talkers they were not being understood they increased their speech level and decreased their rate (but did not become more intelligible). It would appear that speaking rate does not always decrease as speech level increases. This is true only when speech level increases because of combatting noise or unintelligibility. When asked to increase speech level in a quiet environment and with no communications to maintain, the rate and level apparently increase concomitantly. This lack of direct relationship between speech level and rate may be the reason why in this study rate decreases with increasing ambient noise but has a tendency to increase when additional communicators are talking (unless too many are talking).

[8] K. D. Kryter, J. Acoust. Soc. Am. 18, 413 (1946).
[9] T. D. Hanley and M. D. Steer, J. Speech Hearing Disorders 14, 363 (1949).
[10] T. S. Korn, J. Acoust. Soc. Am. 26, 793 (1954).
[11] J. M. Pickett, J. Acoust. Soc. Am. 30, 278 (1958).
[12] T. D. Hanley and M. D. Steer, "Effect of Level of Distracting Noise upon Speaking Rate, Duration, and Intensity," Spec. Devices Center Tech. Rept. No. SDC 104-2-14 (June 1949).
[13] J. W. Black, "A Relationship among Fundamental Frequency, Vocal Sound Pressure, and Rate of Speaking," Ohio State University Research Foundation and U. S. Naval School of Aviation Medicine, Rept. 77 (August 1958).
[14] R. W. Peters, "Changes in Voice Intelligibility, Sound Pressure Level of Response, and Duration of Response as a Function of the Speaker's Being Repeatedly Informed that He is not Being Understood by His Listeners," U. S. Naval School of Aviation Medicine Progr. Rept. No. 50 (May 1955).

CONCLUSIONS

If a communicating pair has direct spoken feedback and is required to maintain relatively error-free communications, then for each increase of 10 dB of thermal noise they raise their speech level 5 dB. They also raise their speech level 5 dB if the number of similarly instructed communicators around them successively doubles. Their rate of word delivery decreases with increasing noise but tends to increase with increasing people (until too many people are added).

ACUSTICA
S. HIRZEL VERLAG · STUTTGART
Vol. 13 | 1963 | No. 5

THE EFFECT OF TALKER-LISTENER ANGLE ON WORD INTELLIGIBILITY

by P. O. Thompson and J. C. Webster
U.S. Navy Electronics Laboratory, San Diego, California

Summary

A series of experiments were conducted in a sound-treated studio to study the effects of talker angle and listener angle on speech intelligibility. Background noise was introduced through two loudspeakers to control the general intelligibility level in the vicinity of the listeners. The noise also served to neutralize the effects of speech reflections in the room. Seventy word-tests were run in three sub-experiments. The talker read a list of fifty C-V-C words for each test. The results indicated that speech intelligibility varies more with listener angle than with talker angle, at least within 45° on either side of the talker. However, the relationship between intelligibility and listening angle may have been influenced by the directionality of the noise source relative to the talker and the listeners.

Zusammenfassung

Eine Reihe von Untersuchungen wurden in einem schallgedämpften Tonstudio durchgeführt, um die Einflüsse von Sprecherwinkel und Hörerwinkel auf die Sprachverständlichkeit zu untersuchen. Um den durchschnittlichen Verständlichkeitspegel in der Nähe der Hörer zu regulieren, wurde Hintergrundgeräusch mittels zweier Lautsprecher in das Tonstudio eingeführt. Dieses Geräusch neutralisierte auch den Einfluß von räumlichen Sprachrückstrahlungen. Siebzig Wortproben wurden in drei Nebenuntersuchungen durchgeführt, in welchen der Sprecher jeweils eine Reihe von fünfzig einsilbigen Wörtern für jede Probe las.

Die Ergebnisse zeigten, daß die Sprachverständlichkeit einzelner Wörter mehr von dem Hörerwinkel als von dem Sprecherwinkel abhängig war, zumindest innerhalb von 45° beiderseitig des Sprechers. Dennoch ist es möglich, daß das Verhältnis von Verständlichkeit zu Hörerwinkel durch die Richtung der Rauschquelle bezogen auf Sprecher und Hörer hätte beeinflußt werden können.

Sommaire

On a procédé à une série d'expériences dans un studio acoustique pour étudier les effets de l'angle d'écoute et de l'angle de la parole sur l'intelligibilité. Un bruit de fond était envoyé sur deux haut-parleurs de façon à pouvoir contrôler le niveau général d'intelligibilité dans le voisinage des écouteurs; le bruit servait également à neutraliser les effets de réflexion dans la salle. Dans trois expériences primaires, on se servit de soixante dix mots «test». Pour chaque essai, le «parleur» lisait cinquante mots C.V.C. Les résultats ont montré que l'intelligibilité variait plus avec l'angle de direction de «l'écouteur» que celui du «parleur», pour des variations d'angle du «parleur» allant jusqu'à 45° de chaque côté. Quoiqu'il en soit, la relation entre l'intelligibilité et l'angle d'écoute peut avoir été influencée par la directivité de la source sonore comparativement au «parleur» et à «l'écouteur».

1. Introduction

Recently Chalupova and Slavik [1], working with listeners who were stationed in a rank and file formation, found that those who were directly in front of the talker often scored lower than those who were slightly to the side, even out to an angle of 45°. Although the cause of this effect could have been the listening angle (the relative direction the listener was facing), the experimenters suggested it might have been due to the directional characteristics of the speech or, in other words, the talker angle. Since they did not independently control either variable, they could not specify the responsible agent.

Because of the curiosity stirred up by this question of a possible gain in intelligibility 15° to 45° off the speech axis, an effort was made to coordinate work on this question with a research program aimed primarily at the effects of noise characteristics and talker distance on word intelligibility and vocal effort. The background noises required by the program would mask the indirect, reflected components of the speech signal and would thereby make possible a study of the directional characteristics of the direct component.

2. Procedure

The major variables in the experimental program, into which the listener and talker angle sub-variables were woven, were word intelligibility vs. distance between talker and listener, and type of ambient noise. By counterbalancing the location of the listeners around the talker, the direction the listeners faced, and the direction the talker faced, it was possible to get the major data of distance and noise type and still extract data on listener and talker angle.

Three experiments were conducted: The first dealing primarily with the effects of distance between talker and listener (1.8, 3.6, and 0.9 m), the second primarily on talker and listener angle, and the third on the type of background noise used. All three experiments were conducted in a sound-treated studio that was especially designed for recording speech. Cylindrical surfaces were purposely designed into the room to achieve uniform diffusion of reflected sound. A plan view of the test room, including the layout of noise sources (loudspeakers), and the general positions of talker and listeners is presented in Fig. 1. The loudspeakers were positioned such that the masking noise gave the most uniform coverage over all listening positions.

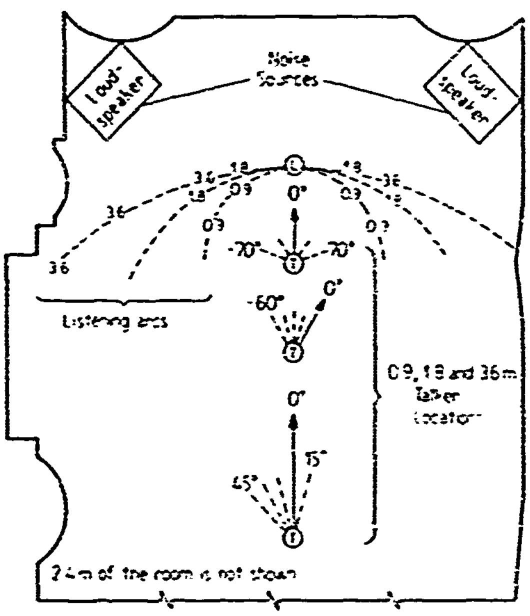


Fig. 1. Plan view of the arrangement of the talker locations and listener locations (the sets of four 0.9's, 1.8's and 3.6's near the top in each case in combination with the "L", which was the listener location common to all three arrangements). The angular limits of the three listening formations (dashed) and the direction the talker generally faced (solid) are indicated at the talker locations. In the third experiment the whole 1.8 m talker-listener complex was rotated 15° either to the left or to the right for some of the tests.

Rhyme word lists were used for all tests [2]. Each list is composed of 50 C-V-C words. Rhyme words retain the same vowels and final consonants from list to list, but the initial consonants change from list to list.

2.1. Experiment I

In this experiment the distance between the talker and the listeners was systematically varied from 1.8 to 3.6, to 0.9 m. At all three distances, (1) the listeners faced the talker half the time and 180° away from him half the time, (2) each listener advanced counterclockwise to the next chair after each test word list, and (3) jet aircraft noise provided the masking background for listening.

At the 0.9 m distance the talker faced the center listener. The remaining four listeners were spaced on an arc symmetrically around the center listener at ±35° and ±70° (see Fig. 1). The jet noise sound-pressure-level (SPL) varied from 64 through 76, 88, 100, to 106 dB(C) for each pair of word lists (half with listeners facing the talker and half with listeners facing away from the talker).

At the 1.8 m distance the talker faced the most clockwise listener position. The remaining four listeners were spaced 15° apart along the arc of 1.8 m radius (see Fig. 1). The jet noise levels were 64, 76, 88, 94, and 100 dB(C).

At the 3.6 m distance the talker again faced the central listening position. But the other listeners were not located symmetrically around the central position but at ±15°, −30°, and −45° (see Fig. 1). The jet noise levels were 64, 70, 76, 82, and 88 dB(C).

Experiment I included two listener angles (facing): 0° and 180°, and many talker angles 0° and angles in steps of 15° to 60°, and in steps of 35° to 70°. The off-centering of the talker (1.8 m distance) and the listeners (3.6 m distance) was an attempt to balance out the geometry of the room-loudspeaker-chair locations and arrive at true talker-angle functions. The different noise levels at different talker-listener distances was an attempt to get roughly equivalent test scores for the various distances.

2.2. Experiment II

Experiment II was divided into two parts of five tests each. For both parts the noise background was 88 dB jet noise and the talker-listener distance was 1.8 m. In the first part the listening angle was varied from one test to the next, from 0° through 30°, 60°, and 90°, to 180°. The talker faced the listener farthest clockwise (0° at the 1.8 m position in Fig. 1) for all five tests. Each listener moved one place counterclockwise after every test.

In the second part the listeners did not change positions at all and always faced the talker. The talker faced the listener farthest clockwise for the first test and on each succeeding test he rotated 15 to the left to face a different listener. In this way talker angle was less dependent upon chair location.

2.3. Experiment III

In experiment III a set of five tests was run for each of six different background noises at a talker-listener distance of 1.8 m. However, for the purposes of this report the difference in noise background is irrelevant and will not be discussed further.

In addition to the five listener positions shown in the 1.8 m arc of Fig. 1, an additional one was added on each side, 15° beyond the outside ones indicated in the figure. For two sets of five tests each, the listeners sat in the five chairs farthest clockwise, for another two sets they sat in the five regular, interior positions, and for the other two sets they sat in the five positions farthest counterclockwise. The use of these seven listener positions made talker angle less dependent upon specific single positions. In each pair of sets the talker faced the listener farthest to his left for one set and the listener farthest to his right for the other set. In every set of five tests the listeners faced the talker on the first test and then 30°, 60°, 90°, and 180°, respectively, from him on succeeding tests. After every test the listeners also exchanged seats in a counterclockwise direction.

3. Results

3.1. Experiment I

The results, corrected to give the data for each talker-listener distance the same mean, are plotted in Fig. 2 with talker angle for various distances as the parameter. The 0.9 m data show a drop in intelligibility as the talker angle increased from zero, while the longer distance data show an increase, at least out to 30°. The loss from listening with backs to the talker was not as great as might be expected. In one 2-test comparison it was 13.7 per cent but otherwise it ranged from 0 to 7 per cent.

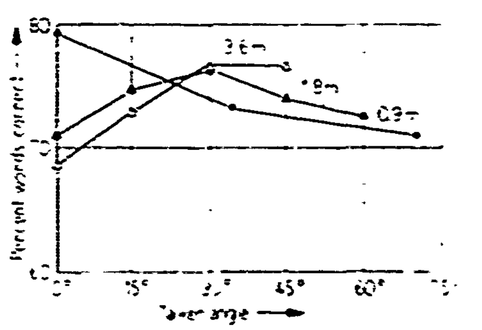


Fig. 2. Word intelligibility results of experiment I as a function of talker angle (the angle between the direction the talker faced and the location of the listener). The parameter is the distance between the talker and the listeners. The talker was immobile, so that talker angle was not independent of the locations of the listener's chairs. The listeners moved one position counterclockwise after every list, facing toward the talker (listening angle 0°) on half the lists and 180° away on the other half.

3.2. Experiment II

Fig. 3 shows the results of experiment II. The listeners 15° and 30° from the direction the talker faced scored higher, on the average, than the listeners directly ahead at 0°. Further, the listeners scored higher on the average, when they did not face the talker directly but instead faced away from him by 30° and 60°. 90° listening was about equal to 0° listening, and 180° listening was substantially worse.

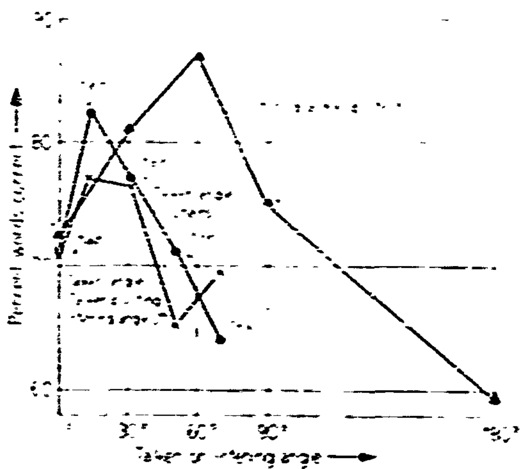


Fig. 3. Results of experiment II. The listening angle results (▲) are averaged over all talker angels and show the effect of rotating at a fixed listening location through 30°, 60°, 90°, and 180°. The talker angle results are of two types: (1) a particular talker angle represented by a particular listener chair or location with talker immobile and (2) talker angle varied by rotation of the talker between lists, facing a different listener location on every list and making a specific talker angle independent of a specific listener location in the room. In this second type each listener remained in a particular chair throughout.

3.3. Experiment III

Fig. 4 shows a gain in intelligibility at oblique listening angles similar to but less than the gains shown in the experiment II data (Fig. 3), and, in contrast, in this experiment the mean score for listeners directly in front of the talker is as high as for any other angle. (The 0° and 30° mean scores are equivalent for these six sets of lists.)

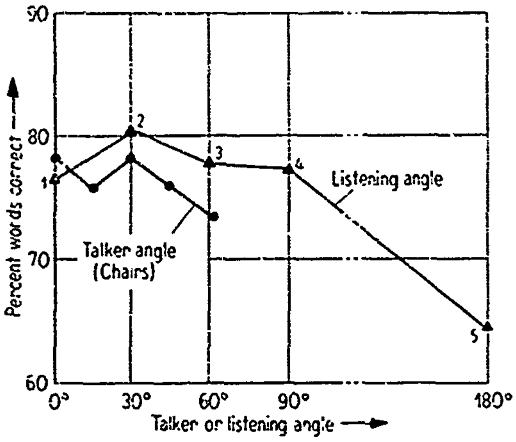


Fig. 4. Results of experiment III. The talker angle data is for the following test conditions: the talker faced alternately one end of the listening formation or the other for each group of five lists. but was immobile within such a group of lists, while the listeners (between lists) were exchanging chairs. The whole listening arrangement was shifted left 15° for 1/3 of the lists and right 15° for another 1/3 of the lists by using one additional chair on each end of the regular five-chair formation. The listening angle data was obtained by rotating the listeners through 30°, 60°, 90°, and 180° from list to list.

4. Discussion

In experiments II and III (Figs. 3 and 4) where listening angles were varied extensively, listening angles of 30°, 60°, and 90° resulted in better average scores than 0°. The results on the effect of talker angle in the three experiments showed a number of cases of highest intelligibility directly ahead of the talker as well as a number of cases in which it was highest at some angle within the range of 15° to 45° to the side. In the first two experiments, out of six sets of five tests each (excluding the 3.6 m data with its poorer uniformity of noise in the listening area), four turned out in favor of some oblique talker angle. In experiment III, however, the result was that in three of six sets of tests, 0° talker angle resulted in a higher mean.

An "Analysis of Variance" treatment of the first series of tests in experiment II (the only set subjected to such analysis) classified the differences found on the listening-angle variable as "very significant." However, the effect of talker angle, which in this set favored 15° and 30° quite substantially, was found to be "non-significant" by this analysis.

The listening-angle effect in experiment III was somewhat less but essentially the same.

4.1. *Talker-angle effects*

What generalizations can be made concerning the effect of talker angle? Starting with the experiment I data for the 0.9 m talker-listener distance (Fig. 2) (which because of highest ratio of direct speech component to indirect components may be the most valid) the 0° talker angle yielded the highest intelligibility. This holds true regardless of whether the listener was facing frontwards or backwards. In the other experiment I data shown in Fig. 1, oblique angles yielded highest intelligibility. In experiment II (Fig. 3) regardless of the experimental procedure, small oblique angles yielded the highest intelligibility. In experiment III (Fig. 4) 0° talker angle yielded the highest intelligibility again. Experiment III should be heavily weighted in consideration because it represents six sets of five tests each and the effects of angle of talker-facing and orientation in the room were fairly well neutralized by counterbalancing. In this experiment 0° talker angle averaged highest in intelligibility with the 30° intelligibility essentially the same.

In almost every case considered in the three experiments the result was different, sometimes favoring one angle and sometimes another. Considering the total 0.9 and 1.8 m data by sets (groups of five tests), in four of the twelve sets intelligibility was highest at 0°; in another three sets it was highest at 15°; in three sets it was highest at 45°; and in two sets it was highest at 30°. On the basis of these results it appears that intelligibility is best in front of the talker but that intelligibility is essentially equivalent over a broad arc in front of him.

As another check on the effect of talker angle the results of all ten sets of 1.3 m data were averaged and the means for talker angles of from 0° to 60° were 75, 76, 77, 74, and 72 per cent respectively. On the average, there was a trend of higher intelligibility at slight angles off the speech axis, amounting to 2 per cent at 30°. This is tempered, however, by the fact that for the average of six of the ten component sets (all of experiment III), the 0°- and 30°-mean scores were equivalent and in three of the six sets of tests 0° talker angle yielded the highest score. These experiment III scores are corrected for over-all intelligibility difference from chair to chair. Before correction the intelligibility advantage of 0° was prominent. Corrections reduced this by 2.5 per cent.

4.2. *Noise direction problem*

The noise SPL varied no more than ± 1 dB within each seating arrangement, except in the case of the two most counterclockwise listening locations in the 3.6 m data. The superior intelligibility at talker angles of 30° and 45° in the 3.6 m data can be attributed to this factor. The next question is whether noise direction characteristics were equivalent from one listener location to another. Subjectively to the listener the noise source was the loudspeaker closest to him. except that on the mid-line between the loudspeakers the apparent source was a point midway between them. A survey of the spatial arrangements (Fig. 1) reveals that, aside from the obvious inequality of both noise level and direction characteristics in the two 3.6 m locations mentioned above, the other listener locations were essentially equivalent in these respects. As regards noise direction characteristics, for example, the apparent source of noise for all five listeners was quite close to 180° from the direction of the talker. This similarity supports confidence in the validity of the talker angle results, even though these talker angles were not wholly independent of specific locations in the room.

A final question needing attention is the possible effects of the noise direction characteristics on the listening angle variable. According to the findings of KOCH [3], when speech and noise are separated by exactly 180° (re. the listener's location), listening angles of 0° and 180° worsen by up to 12 dB the minimum detectable (binaural) threshold for speech, as compared to listening angles such as 45°, 90°, and 135°.

It seems logical that the directivity effect upon intelligibility would be very similar to the effect upon threshold and it is quite possible that noise direction characteristics related to listening angles of 0° and 180° in this study handicapped intelligibility in these cases. However, there is a possibility that KOCH's findings should not be applied to this study because he used an anechoic room, while the present study was conducted in a fairly reflective room. (The room volume of this voice-recording studio is about 100 m³ and the reverberation time is about 0.5 s above, and about 0.3 s below 1 kc/s.)

Using very similar methods to those of KOCH, HIRSH [4] found slightly less effect (9 dB) for the same conditions, and, when a "highly reverberant" room was used, the effect was almost completely destroyed. In the reflective room the average difference in thresholds he found between "Front-back" and "Back-front" versus "Right-left" (0° and 180° versus 90°) was only 2 dB.

The room HIRSH used was probably much more reverberant than the one used in this study and the effect in the present data would be more than in the very reverberant room but less than in the anechoic room. An agent that might have reduced this artifact in the present study is the fact that the noise was not exactly 180° for all listening positions relative to the talker, thereby giving the listener some extra benefit from binaural localization cues (lateral separation of speech from noise). Both HIRSH and KOCH rigidly controlled source angles, but only at 90° intervals in the case of HIRSH and only at 45° intervals in the case of KOCH. It is possible that 5° or 10° offset of the apparent noise source could have drastically reduced the effect, adding to the neutralization caused by the reverberation characteristics of the room.

The 2 dB effect HIRSH found remaining in the reverberant case is enough to account for the roughly 5 per cent (words correct) improvement at listening angles of 30° and 60° found in this experiment. Therefore, whether the effect found here is due to pure listening angle (dependent only upon the angular relationship of talker and listener) or is an artifact of the directional characteristic of the noise cannot be answered.

5. Summary and conclusions

A talker read 70 lists of 50 monosyllabic words each to five listeners in a room designed for the recording of speech. The listening was done in a noise background. The variables studied were talker angle, listening angle, and their effect upon speech intelligibility. The angular relationship among talker, listeners, noise sources, and room were varied extensively to isolate the effects of talker angle and listening angle from most of the other factors.

The results support a view that talker angles (at least) within the range between 0° and ±45° do not have any differential effect on intelligibility. The result of varying listening angle away from 0° to 30° and 60° was higher intelligibility. This result, however, is open to question in the light of the findings of HIRSH and KOCH, which suggest that with the configuration of speech source and noise source used in the present experiments, binaural discrimination factors substantially handicap intelligibility at listening angles of 0° (and 180°), at least as compared to angles from 45° to 135°.

Only further experiments under free-field conditions can give adequate appraisal of the effect of (binaural) listening angle on intelligibility. If environmental noise is introduced to control the general level of intelligibility, it must be controlled in such ways as to neutralize unwanted binaural discrimination cues. Or the experiments could be conducted in open air so that the direction of noise incidence would have a quasi-random characteristic.

(Received Nov. 16th, 1962.)

References

[1] CHALUPOVA, V. and SLAVIK, J. B., Über die Kurven gleicher Sprachverständlichkeit. Preprint H 47 of a paper given at the Fourth International Congress of Acoustics in Copenhagen, August 1962.
[2] FAIRBANKS, G., Test of Phonemic Differentiation; the Rhyme Test. J. Acoust. Soc. Amer. **30** [1958], 596.
[3] KOCH, W. E., Binaural localization and masking. J. Acoust. Soc. Amer. **22** [1950], 801.
[4] HIRSH, I. J., The relation between localization and intelligibility. J. Acoust. Soc. Amer **22** [1950], 196.

ACUSTICA

S. HIRZEL VERLAG · STUTTGART

Vol. 14 1964 No. 1

THE EFFECT OF TALKER-LISTENER ANGLE ON WORD INTELLIGIBILITY

II. In an open field

by P. O. Thompson and J. C. Webster

THE EFFECT OF TALKER-LISTENER ANGLE ON WORD INTELLIGIBILITY

II In an open field

by P. O. Thompson and J. C. Webster

U. S. Navy Electronics Laboratory, San Diego, California

Summary

Two experiments were conducted to study the effects on speech intelligibility of talker angle and listener angle and the distance between talker and listener under free-field conditions. Some findings were that (1) speech intelligibility falls off with distance in a manner consistent with a 3 dB per distance doubled fall-off in speech sound pressure level; (2) the intelligibility in a broad arc from $-45°$ to $+45°$ in front of the talker was essentially equal; (3) the effect of turning the listener $15°$ to $75°$ away from the talker was a mean gain of about 4% or a gain equivalent to about 3.5 m in distance; (4) the observed directional aspect of intelligibility agreed quite well with SPL measurements made by other investigators around the heads of a model and a human; and (5) the effects of distance and wind in this study were in general agreement with the results of previous studies.

Zusammenfassung

Zwei Versuche wurden durchgeführt, um den Einfluß des Sprecherwinkels, des Hörerwinkels und des Abstandes zwischen Sprecher und Hörer auf die Sprachverständlichkeit im freien Schallfeld zu untersuchen. Unter anderem wurden folgende Beobachtungen gemacht: (1) Die Sprachverständlichkeit fällt mit der Entfernung in einer Weise, die mit einer Verringerung des Schalldruckpegels um jeweils 3 dB bei Verdoppelung des Abstandes erklärt werden kann; (2) die Verständlichkeit war über einen breiten Bereich von etwa $-45°$ bis $+45°$ vor dem Sprecher im wesentlichen konstant; (3) Drehung des Hörers um $15°$ bis $75°$ vom Sprecher fort brachte eine Verbesserung der Verständlichkeit um etwa 4%, was einer Verringerung der Entfernung um etwa 3.5 m entspricht; (4) die beobachtete Richtungsabhängigkeit der Verständlichkeit stimmt gut überein mit Schalldruckpegelmessungen anderer Autoren an Modellen und Personen; (5) der Einfluß von Entfernung und Wind war bei dieser Untersuchung weitgehend der gleiche wie in früheren Untersuchungen.

Sommaire

On a fait deux expériences pour étudier les effets sur l'intelligibilité de la parole, de l'angle d'écoute, de l'angle de parole, et de la distance entre auditeur et «speaker» dans les conditions de champ libre. Quelques résultats ont montré que:

1) l'intelligibilité diminue avec la distance, ce qui se traduit qualitativement par un affaiblissement de 3 dB du niveau sonore moyen par distance doublée.
2) l'intelligibilité reste la même pour un angle d'ouverture allant de $-45°$ à $+45°$ de part et d'autre du «speaker».
3) lorsque l'on tourne l'auditeur de $15°$ à $75°$ du «speaker», le gain moyen est d'environ 4%, ce qui correspond à un gain en distance de 3.5 m.
4) l'effet directionnel de l'intelligibilité de la parole que nous avons observé est parfaitement en accord avec les mesures SPL faites par d'autres chercheurs opérant avec la tête d'un mannequin et d'un homme.
5) les effets de distance et de vent dans cette étude sont en général en accord avec les résultats des études précédentes.

1. Introduction

CHALUPOVA and SLAVIK [1] recently studied intelligibility as a function of the angular relationship between the talker and listeners in a rectangular formation of rows and columns in an open field. Their most interesting finding was that the intelligibility on the speech axis was lower than it was 10° to 45° off the axis. Their experimental design did not allow them to determine whether this was caused by directional characteristics of the voice or the directional characteristics of the external ear. Other studies [2], [3] have not shown a lowering of intelligibility on the speech axis, but perhaps this was because they were not as intensive in angular coverage, number of listeners, and number of replications.

In order to check the CHALUPOVA and SLAVIK results two experiments were conducted on an open field at the U.S. Naval Training Center in San Diego. In both experiments Navy recruits were the listeners, and the parade grounds, Preble Field, was the open field. Preble Field is 305 m long by 137 m wide and is surfaced with asphalt paving.

To study the dependence upon the listener's location relative to the talker's mouth, in experiment I the talker rotated through a series of angles relative to the listeners. To study the dependence upon the direction the listener was facing relative to the talker's location, the listeners in both experiments were divided so that at any time half of them faced the talker and the others faced away from him by small angles.

2. Procedure

2.1 General

The intelligibility tests used were the FAIRBANKS Rhyme Tests [4]. These are lists of 50 monosyllabic words (consonant-vowel-consonant) such as "1. cop, 2. wire, ----, 50. ten". There are five basic tests using words that not only rhyme from test to test, but are spelled the same except for the initial consonant. For instance, a second test might be "1. pop, 2. sire, ---- 50. den", or "1. top, 2. wire, ----, 50. pen", etc. The answer sheets show the vowel and final consonant of each word (thus, "1. –op, 2. –ire, ----, 50. –en").

Listeners were instructed to fill in the initial consonant of each word as it was read, even if they had to guess. In each experiment a different rhyme word order was used for every test conducted. Fourteen 50-word lists were used in each experiment. The listeners were thoroughly instructed in the testing procedure and were given a practice test.

The talker for these experiments was the first author, who has a resonant, moderately pitched voice, is somewhat above average in articulation, and is very experienced as a talker for intelligibility tests. The talker monitored his speech level on a small sound-level meter mounted 0.5 m in front of him at chest height. He spoke with raised voice, maintaining an average 85 dB sound pressure level (SPL) on voice peaks. This corresponds to an rms SPL of 75 dB at 1 m. The word "write" was read just ahead of each test word, both to act as a "carrier" for the test word and to be used as a calibration word for level monitoring. One test word was read every 2 s to the beat of an electronic interval timer. When extraneous noises were of such a nature as to interfere with speech reception, the talker stopped until he considered the noise level within acceptable limits again. The noises that caused occasional temporary stoppages were from aircraft flying in the vicinity. A microphone worn around

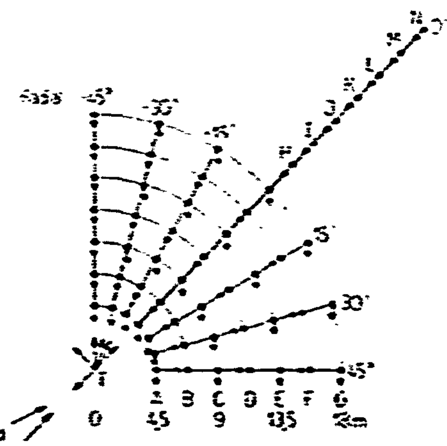


Fig. 1. Talker-listener locations and facings for experiments I and II. In experiment I, 42 listeners sat at intersections of arcs A through F and 7 radials spaced at 15° intervals. In experiment II listeners also sat on arc G. In experiment I, the talker (T) faced the center (0°) radial on tests 1, 2, 13, and 14, the −15° radial on tests 3 and 4, the −30° radial on tests 5 and 6, the −45° radial on tests 7 and 8, −90° on tests 9 and 10, and 180° on tests 11 and 12. In experiment II the talker always faced the 0° radial. In both experiments listeners in arcs B, D, and F faced the +45° radial on all odd-numbered tests, while listeners in arcs A, C, and E (and G in II) faced the talker location. The listeners reversed their direction of facing on even-numbered tests, i.e., those who had faced the +45° radial faced the talker, and those who had faced the talker faced the +45° radial. Note that on the −45° radial, facing the talker was equivalent to facing the +45° radial. In both experiments the wind direction varied between west and west-northwest; in experiment I wind velocity varied between 27 to 35 km/h and in experiment II wind velocity varied between 0 and 9 km/h.

the neck by the talker was used to record the test sessions for later review of speech and noise levels.

As illustrated in Fig. 1. the listeners sat in concentric arcs 2.3 m apart, starting 4.5 m from the talker. Listeners sat behind one another on 7 radials spaced every 15° from −45° to +45° relative to the talker's standard direction of facing. By sitting on the ground each listener had an unobstructed pathway between his ears and the talker's mouth. By sitting cross-legged it was easy to maintain a rigid and accurate listening angle.

In both experiments listening angle was varied such that listening angles of either 0°; or 0°, 15°, 30°, 45°, 60°, 75°, and 90° resulted for the seven radials from left to right, respectively.

The talker's angle, the angle the talker faced with respect to the listener position, was varied in experiment I but was constant in experiment II.

2.2 *Experiment I*

In experiment I, forty-two listeners were used. Each listener remained in the same location throughout all tests, although, as illustrated in Fig. 1. each rotated systematically between two different facings.

The talker angle was varied as detailed in Fig. 1.

Experiment I was conducted in the afternoon and the prevailing Westerly wind interfered substantially with the intelligibility of the speech.

2.3 *Experiment II*

In experiment II, a second group of 49 recruits was seated in 7 arcs and 7 radials. These 49 listeners were selected from the 60-odd regular members of the available recruit company on the basis of the practice test scores. In order to reduce the effect of listeners on location, each listener was shifted diagonally back one arc and over one radial after each pair of tests. In this manner each listener was in each arc and in each radial for one pair of tests. The alternation in the listening angles between the odd and even tests in each pair of tests was executed just as in the first experiment, and the other testing details were the same.

The experiment II session was held in the morning before the prevailing winds could interfere and the wind velocity was only 0 to 9 km/h.

In experiment II the talker always faced the center of the group, the 0° radial.

3. Results

3.1 *Intelligibility vs distance*

In Fig. 2 are plotted the results of experiments I and II in terms of words correct vs distance from the talker. The fact that better scores were obtained on experiment II than on I could be due to sampling error but more probably the different scores are due to differences in wind velocity or experimental design.

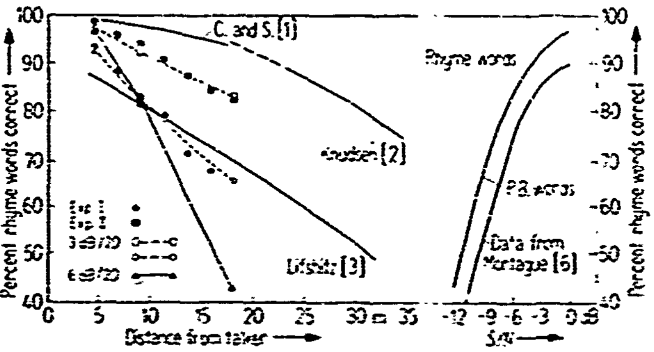


Fig. 2. On the left, per cent of Rhyme words correct vs distance from talker. Each of the experimental points (solid circles and squares) represents an average score for 7 listeners and 14 tests of 50 words each. All talker-angle and listener facing-angle data are combined. Also shown are intelligibility data vs distance from three other studies. The open symbols and dotted lines represent intelligibility scores predicated on a 3 dB drop in level for each distance doubled. The triangles represent a 6 dB per-distance-doubled drop in intelligibility. On the right is per cent words correct vs speech-to-noise differential in dB for Rhyme and PB words. This is the data used for the 3 and 6 dB per-distance-doubled data plotted on the left.

Concerning the wind: In experimental I the gusts of wind averaged 27 km/h and peaked at 35 km/h. In experiment II the wind velocity varied from 0 to 9 km/h. Hayes and Cudworth [5] found that when a listener faced a 33 km/h wind, his thresholds for tones in the frequency range of speech were shifted by 40 to 45 dB. By turning away 45° and 90° this shift was reduced to 30 and 15 dB, respectively. For a wind of 18.5 km/h they found that the threshold shift was only about half as great. Since the winds came in gusts, the results of experiment I are not as different from II as the Hayes and Cudworth data [5] might suggest, but for oserrers on the spot at the time there was no question but that the excessive winds during experiment I reduced word intelligibility scores.

In the data reported by Knudsen [2], a wind of 33 to 41 km/h reduced the distance for 75% intelligibility from 43 m to 26 m. In the present experiments the 75% coverage was reduced from 25 m to about 15 m, a reduction of the same proportion. This would suggest that wind interference was the near-exclusive cause of the general intelligibility difference between experiments I and II.

The effect that the experimental design might have had on depressing the scores in experiment I versus II will be discussed in the section of effects of talker angle.

Plotted as solid lines (or extrapolated dashed lines) in Fig. 2 are the results of other speech intelligibility tests in open spaces reported by Chalupova and Slavik [1], Knudsen [2], and Lifshitz [3]. Except for experiment I, the windy one, the rate at which intelligibility falls off with distance is fairly consistent. By utilizing the information plotted on the right of Fig. 2, it is possible to check whether the rate of decrease in intelligibility with distance follows a decrease in sound pressure level of 6 dB or 3 dB per distance doubled. The Montague [6] data concerns the relation between word intelligibility and speech-to-noise ratio for Rhyme words (used in experiments I and II) and for PB words (extensively used in other tests of this type). The open symbols on the left of Fig. 2 were derived from the data on the right of Fig. 2 to show that decrease in word score per distance doubled is very exactly predicted by successive decreases in speech level of 3 dB. The triangles on the left show that the word scores with distance doubled do not fall off as fast as would be predicted by successive decreases in speech level of 6 dB.

This implies that the physical laws of cylindrical waves in open space are being followed and that the physical laws of spherical waves are not.

These results bear on one of the assumptions inherent in Beranek's formulation of Speech Interference Levels [7], [8]. The assumption is that to retain equal speech intelligibility the acceptable noise level must decrease 6 dB per distance doubled for distances up to 4 m. The results of the present experiments imply that beyond 4.5 m, even in open air, the acceptable noise level need decrease only 3 dB, not 6 dB, per distance doubled. This statement assumes of course that speech level is always really a speech-to-noise differential and that a decreasing speech level is exactly equivalent in dB to an increasing noise level.

3.2 *Intelligibility vs talker angle*

In experiment I the data on angle of listener facing are subject to large interactions involving wind direction. Those listeners who faced most nearly into the wind when they faced the talker showed larger effects due to differences in listener facing than those listeners who were not facing into the wind when they faced the talker. Experiment I was not designed to take account of the wind and the listener angle data have a wind bias that cannot be isolated.

Information on talker angle, however, is not so dependent on the wind because the talker rotated through many angles and, as he did, the relative position between him and the listeners was somewhat balanced with respect to wind direction. That is, as the talker faced the center of the group, those listeners slightly to his right faced the wind, but as the talker rotated to 90° these slightly to his right were at right angles to the wind. Therefore the remaining discussion of experiment I will be limited to the results on talker angle.

Fig. 3 shows a polar plot of word intelligibility as a function of talker angle; that is, the word intelligibility as a function of where the listener sat with respect to the direction the talker faced. The direction the listener faced with respect to the talker is not considered.

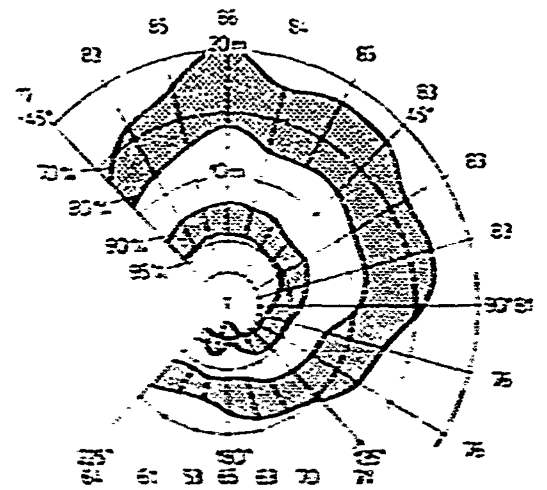


Fig. 3. Equal intelligibility contours vs angle from talker for lumped listener-facing angles. Contours are derived from a breakdown of the talker angle information from the experiment I data in Fig. 2. The numbers on the periphery are mean words correct for each talker angle.

The numbers in parentheses on the plot show the average percentage of words correctly heard by all listeners who sat on radials that were at one time or another from −45° to +225° with respect to the talker's mouth. These data show that listeners heard distinctly better when in front as compared to behind. The low value of 77 at −45° is based on only one-third the number of listeners as the value at +45°; similarly, the data at −30° represents only three-fifths the number of cases as at +30°. Therefore a truer picture of the −45° to 0° results would probably be the mirror image of the 0° to +45° results, since the positive angle results are based on the results of many more listeners.

The data are presented as contours of equal intelligibility so that the results can be more easily compared to the Chalupova and Slavik [1] results. To draw equal intelligibility contours, all the data at each talker angle were first plotted as a function of distance (arc). Curves were then visually fitted to these data points. The intersections of these curves with discrete intelligibility levels of, say, 95, 90, 85, 80%, were then read off and plotted in Fig. 3.

The equal intelligibility data, just as the average of all the data, show that listeners in front of the talker got decidedly higher scores than listeners behind the talker. Scores from $-45°$ to $+45°$ are roughly equal and only slightly greater than those out to angles of 75°.

3.3 *Intelligibility vs listener facing angle*

Experiment II was designed to get reliable information on angle of listener facing. A time was chosen when the wind velocity was low, the talker always faced the center of the listening group, and after each series of two tests the listeners moved diagonally back one arc and counterclockwise one radial.

Equal intelligibility contours from the results of experiment II are plotted in Fig. 4. The two lines at each contour represent those who always faced the talker (the lower boundary of the shaded area) and those who faced away from the talker by the angle shown on the top arc. These were the same listeners, but half the time they faced the talker and half the time they faced normal to the most clockwise radial. When the listeners faced away from the talker, their scores consistently averaged higher, or, as plotted in Fig. 4, they got the same average score at greater distances from the talker.

The mean intelligibility scores for the listeners, when their listening angles were varied, averaged 87, 90, 91, 93, 92, 92, and 91% for the angles from 0° to 90°, respectively, in 15° increments. This amounts to an advantage for the 30°, 60°, and 90° angles of 4, 5, and 4%, respectively. This compares with an advantage of 13, 14, and 12%, respectively, for the same angles in an investigation by NORDLUND and FRITZELL [9], who worked in the 70% intelligibility region with speech picked up binaurally via a dummy head.

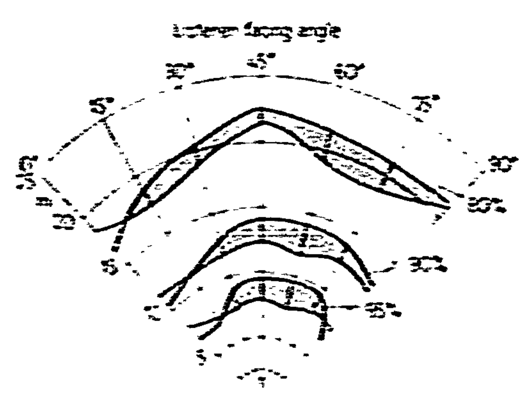


Fig. 4. Equal intelligibility contours derived from the data of experiment II. The talker always faced the center of the group so information on talker angle is read symmetrically around the center radial, from 0° through $\pm 15°$, $\pm 30°$, and $\pm 45°$. The listener facing angle was either 0° (bottom line of shaded area) or the angle designated on the radials reading clockwise from 0° to 90°.

The results agree qualitatively with those of ROBINSON and WHITTLE [10] in their experiments on the effects of angle of reception upon the aural effectiveness of pure tones. In their binaural loudness tests they observed gains over a similar angular range at 2500 and 4000 c/s. Their measurements of SPL at the canal entrance of the ear also show an average gain at these angles, as compared to the case when the source was dead ahead. Their measurements closely correspond to the measurements of WIENER and ROSS [11], whose mean 45° and 90° SPL readings at the entrance to the ear canal averaged 4 or 5 dB better than the 0° SPL readings over the frequency range from 400 to 6000 c/s.

4. Comparison to other data

The data of experiment I and the data on 0° angle of listener facing of experiment II have been combined to obtain three sets of equal-intelligibility contours, plotted in Fig. 5. These contours represent generalized data on talker angle. Also, on the right of Fig. 5 are the polar plot displays of selected contours from CHALUPOVA and SLAVIK [1], and LIFSHITZ [3], and an extrapolated contour based on the data reported by KNUDSEN [2]. On the left of Fig. 5 are some SPL data of DUNN and FARNSWORTH [12] and of FLANAGAN [13]. These data show the directional patterns of sound from a source like a mouth.

Concerning the intelligibility contours shown on the right, it is evident that the present data are far more squat in shape than any of the others. As compared to the other contour shapes, the present data are relatively flat on top (from 0° to $\pm 45°$).

As regards the depression directly in front of the talker, which was observed in the CHALUPOVA and SLAVIK data, this was possibly caused by interference effects of the bodies of listeners between the talker and other listeners and by the slight disadvantage that exists for listeners who face the talker directly as compared to those who face at small angles away from the talker. In the rectangular formation used in their experiment the degree that listeners were in the shadow of other listeners was dependent both on the particular angle from the talker and the distance from the talker. This would cause angular and distance effects on the transmission of the highly directional high frequency components of the speech. This in turn could have systematically affected the reception of consonants in such a way as to account for a large share of the detail in their equal intelligibility contours.

It is quite evident that the four studies do not agree in the magnitude of the percentage of correctly heard words vs distance. Since the present interest is in contour shape, this need not be of too great concern. Talker level, wind velocity, temperature gradients, differences in vocabulary, and even differences in motivation could acount for these discrepancies.

On the left it is interesting to note that the present data agree very well with the SPL of speech measured in the 62 to 12,000 c/s band by DUNN and FARNSWORTH [12]. The FLANAGAN [13] and the 2 to 2.8 kc/s data of DUNN and FARNSWORTH lie about half-way between the shapes of the present data, and LIFSHITZ, and CHALUPOVA and SLAVIK data which have steeper sides from about $\pm 45°$ to $\pm 90°$.

The FLANAGAN curve in Fig. 5 is based on his SPL measurements of a 2 kc/s tone projected through the mouth of a model [13], while the DUNN and FARNSWORTH curves [12] are based on speech SPL measurements on a live talker. In both cases the measurements were made about 30 cm from the source. To make up these curves for Fig. 5, relative SPL measurements from the plotted data of the references cited were converted to relative distance on the basis that, as distance is doubled, SPL is reduced by 6 dB, since both studies showed that at distances close to the source this is the rate at which sound pressure falls off with distance.

(Received April 18th 1963.)

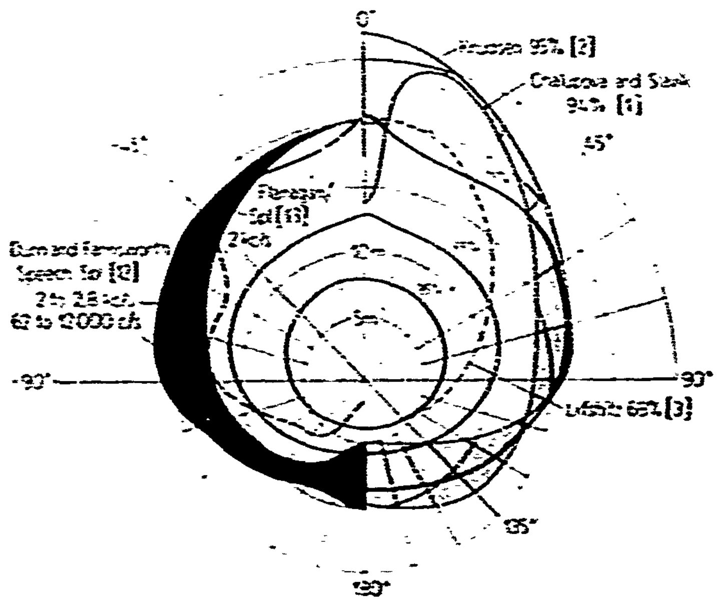


Fig. 5.
Equal intelligibility contours of combined data from experiments I and II and from references [1], [2], and [3]. Contours of equal sound pressure level for (1) broad band and narrow band speech derived from data in reference [12], and (2) tones from an artificial mouth in a model head derived from data in reference [13] are also shown.

References

[1] CHALUPOVA, V. and SLAVIK, J. B., Über die Kurven gleicher Verständlichkeit der Sprache. Fourth International Congress on Acoustics, Copenhagen 1962, H. 47.

[2] KNUDSEN, V. O. Architectural Acoustics. John Wiley & Sons, New York 1932, p. 498.

[3] LIFSHITZ, S., J. Acoust. Soc. Amer. 4 [1932], 112.

[4] FAIRBANKS, G., J. Acoust. Soc. Amer. 30 [1958], 596.

[5] HAYES, J. R. M. and CROSWORTH, A. L., J. Acoust. Soc. Amer. 26 [1954], 254.

[6] MONTAGUE, W. E., Navy Electronics Lab. Rep. 977, June 1960.

[7] BERANEK, L. L., Trans. Amer. Soc. Mech. Engrs. 67 [1947], 97.

[8] BERANEK, L. L., Acoustics. McGraw-Hill, New York 1954, p. 419.

[9] NORDLUND, B. and FRITZELL, B., Acta Otolaryng. 56 [1963], 632.

[10] ROBINSON, D. W. and WHITTLE, L. S., Acustica 10 [1960], 74.

[11] WIENER, F. M. and ROSS, D. A., J. Acoust. Soc. Amer. 18 [1946], 401.

[12] DUNN, H. K. and FARNSWORTH, D. W., J. Acoust. Soc. Amer. 10 [1939], 184.

[13] FLANAGAN, J. L., J. Acoust. Soc. Amer. 32 [1960].

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Navy Electronics Laboratory San Diego, California 92152	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b GROUP None

3. REPORT TITLE

SPEECH INTERFERENCE ASPECTS OF NAVY NOISES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Research and Development Report Jan 1961 to Dec 1964

5. AUTHOR(S) (Last name, first name, initial)

Webster, J. C., and Klumpp, R. G.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
2 September 1965		17

8a. CONTRACT OR GRANT NO. b. PROJECT NO. SF 013 11 01 Task 1357 c. (NEL Problem N3-4) d.	9a. ORIGINATOR'S REPORT NUMBER(S) 1314 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Bureau of Ships Washington, D. C. 20360

13. ABSTRACT

Representative samples of ship, office, and shop noises were analyzed to determine simple methods for rating noise in shipboard spaces in relation to interference with speech communication. For simplicity in speech interference measurement, it is recommended that an average be taken of the octaves centered at 500, 1000, and 2000 cycles per second.

DD FORM 1 JAN 64 1473 0101-807-6800

UNCLASSIFIED
Security Classification

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech - Interference - Measurement Noise - Analysis						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE. Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ______________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ______________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ______________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (*paying for*) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (*TS*), (*S*), (*C*), or (*U*).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

UNCLASSIFIED
Security Classification